Unpacking Learning

Haug, C.C.

2015

Unpacking Learning

Conceptualising and measuring the effects of two policy exercises on climate governance

CONSTANZE HAUG

Leescommissie: prof. dr. M.N.C.Aarts
prof. dr. F.Biermann
prof. dr. R.Hoppe
prof. dr. A.Jordan
prof. dr. I.S.Mayer

Unpacking Learning. Conceptualising and measuring the effects of two policy exercises on climate governance
PhD Thesis, VU University Amsterdam

C.Haug, Amsterdam, 2015
Printed by CPI Woehrmann, Zutphen

The research for this thesis was carried out at the Institute for Environmental Studies (IVM), Vrije Universiteit Amsterdam (http://www.ivm.vu.nl)

VRIJE UNIVERSITEIT

Unpacking Learning

Conceptualising and measuring the effects of two policy exercises on climate governance

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof. dr. V.Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Aard- en Levenswetenschappen
op vrijdag, 20 november 2015 om 13.45 uur
in de aula van de universeit,
De Boelelaan 1105

door

Constanze Carolin Haug

geboren te Fürstenfeldbruck, Duitsland

promotoren: prof. dr. F.G.H.Berkhout
prof. dr. D.Huitema

Table of Contents

Preface

It feels like an eternity ago (and truth to be told, it's not much less) when a young, idealistic graduate chanced upon a vacancy for a junior researcher to support a project by the European Commission on future European climate policy in Amsterdam. What followed was a job interview with very friendly and inspiring future colleagues in beautiful Amsterdam spring weather, with the sunshine glittering over the canals, and a very intuitive, not entirely thought-through decision to 'dig deeper' in climate policy issues than I thought was possible in policy-making, my initial career choice, and to become a researcher instead.

The consequence was a stint in the lowlands much longer than initially envisaged, with high points and low points, lots of things learned, but most of all, beautiful memories and important friendships which will hopefully last for a lifetime. The ADAM project, which initially kept me busy at IVM, proved to be a behemoth among the European research projects, and coordinating remotely across four countries on how to conduct a meta-analysis of European climate policy evaluations turned out more challenging than initially expected. Alongside I went on a quest for a PhD topic that lent itself to being completed next to working on different project deliverables. The idea to focus on games as a means to stimulate learning on climate policy immediately appealed to me, having had a passion for board games and for bringing different people together since childhood. The ADAM project and later on the REDD-ALERT project, which focused on ways to reduce emissions from deforestation in the tropics, provided the setting for conducting the two policy exercises that are at the heart of this book. Designing and organizing them has been both exciting and nerve-wrecking, and I learned a lot in the process. I also learned a lot about things that were more alien to my background, such as social science theory and methods. But perhaps the biggest learning experience was about me and myself. I am not sure these were always things I wanted to know, and as they make risky material for anything published during my lifetime, I will stop navel-gazing right here and turn to those who escorted and supported me.

Turning to the people who were part of this process, my promotors Frans Berkhout and Dave Huitema deserve a huge thank you. Dave was 'in the trenches' with me, acted as an intellectual sparring partner, – the reading clubs on various issues will remain a cherished memory! – encouraged me patiently, and cheered me on to make progress on the PhD rather than getting too distracted with other publications and project work. Dave, I know I strained your patience towards the end of this project, but I hope that, despite the Dutch notoriety for bluntness, you lack the audacity to

disagree with an authority such as Shakespeare: all's well that ends well. Frans, with inimitable diplomatic skills and knack for a birds-eye perspective, was a perfect complement to Dave, encouraging me with constructive high-level feedback, and helping rebuild confidence when necessary. Thanks again to both of you.

Several people generously provided support in the realm of gaming-simulation and without them my policy exercises would not have seen the light of day. Professor Igor Mayer kindly invited me to follow an inspiring course on game design at TU Delft, which provided much of the game design fundamentals required for this thesis. A research stay with Professor Ted Parson at the University of Michigan allowed me to gain important insights into policy exercises, as Ted generously shared his insights and experience with me. Conversations with Ferenc Toth further helped me to understand the ideas behind this method and served to better delineate my choice of topic. Particularly big credit is due to Ivo Wenzler, who took time out of his busy consultant life to help me develop the design for the first exercise and who then did an excellent job facilitating it.

Developing a game rests as much on design skills as on an in-depth understanding of the policy area it seeks to simulate. For the exercise on burden-sharing in EU climate policy, sincere thanks go to the rest of the work package team on EU policy and governance in the ADAM project – Frans Berkhout, Dave Huitema, Harro van Asselt and Eric Massey from IVM, Andy Jordan and Tim Rayner from the University of East Anglia, Johannes Stripple and Roger Hildingsson from Lund, and Suvi Monni from JRC in Ispra. Together, we worked hard to unravel the key 'attention points' and 'dilemmas' in European climate governance. Special thanks in this context to Andy Jordan, the indefatigable editor-in-chief of our opus magnum on the topic, whose skills in 'cutting dead wood' continue to amaze and inspire. Credit is also due to many interviewees who generously shared their expert insights and process knowledge into the issue of burden-sharing / effort-sharing, which helped inform the design of my policy exercise.

The REDD-ALERT project subsequently allowed me to broaden my climate policy focus to tropical deforestation, an issue whose complexities extend far beyond the 'traditional' realm of climate governance. Interesting field trips to Indonesia and Peru offered an initial first-hand glimpse into the realities of deforesting countries, and inspiring discussions with our project partners – amongst others Eric Lambin, Robin Matthews, Patrick Meyfroidt, Peter Minang, Meine van Noordwijk, and Lou Verchot – further helped me get a grasp of the multiple challenges involved in tackling deforestation. A research stay as a YSSP fellow at the International Institute for Applied Systems Analysis in Laxenburg, Austria provided the necessary focus to delve head-on into the literature on this topic – thanks to my IIASA supervisors Michael Obersteiner and Florian Kraxner for hosting me, and to fellow YSSPers for many helpful discussions and feedback. Last but not least, I am grateful to Joyeeta Gupta,

the project leader of our REDD-ALERT work package, who made space in the work package to accommodate my policy exercise, and to Emmy Bergsma who 'covered' for me other aspects of the REDD-ALERT project work. Finally, as in the ADAM exercise on burden-sharing, many interviewees generously gave their time and informed the design of the policy exercise with valuable insights and different perspectives on the topic.

Measuring learning from interactive appraisals was the bold ambition underlying this thesis. A reading club involving Dave, Eric Massey and Gert Becker not only allowed for diverse culinary discoveries but also helped sharpen my initial research focus by provoking a critical reflection on the prevalent concepts and typologies, and the dearth of robust evaluations in this area. A workshop on social learning organised by Stockholm Environment Institute gave a further important boost through helpful discussions on learning with a range of experts. The most important offspring of the workshop was a collaborative project with Ryan Plummer, Julia Baird and Kerrie Pickering at Brock University in Canada. Their interest in our methodology and subsequent application to an interactive appraisal process in Ontario not only led Dave and me to discover Ontario's wineries and the Niagara Falls, but also set in motion a valuable feedback loop that helped advance our typology and evaluation framework and refine the analysis in this thesis. Inside IVM, conversations with Belinda McFadgen, Eric Massey, Stefania Munaretto, Harro van Asselt and Eleftheria Vasileiadou, amongst others, were equally helpful in clarifying and consolidating the ideas put forward in this thesis.

A fundamental precondition for a successful policy game is the right set of participants. I am immensely grateful to the experts some of whom travelled from afar to take part in what to them must have appeared a rather exotic event, and who launched themselves into the experience with great motivation and energy. I am even more grateful for their concept maps, completed surveys and follow-up interviews. Without them, I would have had nothing to analyse and write about. Great thanks are also due to the all the Master's students who participated in the exercises with enthusiasm, and for their completion of the attendant evaluation material. I was surprised time and again how quickly they familiarised themselves with the topics at hand and made them their own. Léa Bigot, Felix von Blücher, Verena Ommer, Hye Young Shin and Timme van Melle played an important part in assisting the preparation and running of the policy exercises, and I thank them for their commitment and efforts.

For a long time, the biggest construction site in my thesis research was the statistical analysis of my data. Tom Louwerse from Leiden University is my hero here, who patiently explained statistical concepts and methods, and provided helpful feedback to ensure that both analysis and reporting of the results were sound. IVM colleagues

Joop de Boer and Eleftheria Vasileiadou were equally very generous with advice on specific questions and issues.

Amidst the distractions of project work, other publications and private life, it wasn't always easy to find the necessary focus for long stretches of concentrated writing. Gigi and Niko & Carolin provided very welcome refuge for two writing exiles in the Austrian mountains and in Berlin, respectively. Generations of office mates on the other hand, from Thijs, Pieter and Stijn, to Derek, Aschalew, Boris, Elko, and finally Nicolien provided pleasant company while working and joined in for the necessary breaks to survive the day. Chats, coffee breaks and borrels with great friends and colleagues, including of course my esteemed paranimfen Eric and Harro, but also Hans, Eleftheria, Aysem, Sander, Agni, Mairon, Hanna, Annick, Stefania, Belinda, Farhad, Oskar, Heather, Jason, Justin, Trond, Emma, Brenden, Marthe, Julia, Philipp G., Patricia, Suzan, Tineke and Els, and many others, were equally a very important part of the excellent working culture at IVM. Last but not least, I would like to thank Frank Biermann, EPA department head, for his continuous support and faith in my abilities throughout my years at IVM. Unfortunately, the Shakespeare quote above does not apply to IVM's fate, but I am grateful that I could experience the inspiring, interdisciplinary culture of the Institute at its best and am sad that future generations of PhD students may not have this privilege.

The good friends I made during my years in Amsterdam will hopefully accompany me for a long time. Julia & Sebastiaan, Masha, Sofia, Eleni, Harro, Eric & Sabina, Ralph & Willeke, Laurens & Silvia, Philipp & Xenia, Marieke & Arjan, Derek – I am so glad we met and will meet again. Our doors in Berlin are always open to you.

Mami & Papi – thank you for giving me the opportunity to go out into the world and explore, and for always supporting me in my ventures. Now having a family of my own, I value the loving base you have provided even more.

Lucia, you'll be glad to hear that there is no more ominous 'book to write' standing in between you and joint visits to the playground or the zoo. I love you and am proud of the confident little girl that you have become – and of course thank you for being my cover artist.

Benni, you will be so relieved to see this book finished. It has tested both of us and I cannot thank you enough for mustering the patience to keep encouraging me until the end. I love you and enjoy (almost) every moment we spend together. I cannot wait to see what life has in store for us next.

.

List of Tables

List of Figures

1. Introduction

1.1 Environmental problems and complexity

> *Environmental protection and restoration are not technically overwhelming – we probably had less of the requisite know-how for putting a craft on the moon in the 1950s than we do for solving major environmental problems today. In our society, environmental problems are democratic dilemmas. The Age of Ecology requires us to build seemingly impossible bridges: somehow hundreds of millions of individual actors must learn the ecological consequences of their behavior, and somehow they must use their knowledge to choose wisely between painful trade-offs (Press, 1994, 1).*

Much has been written about the complexity of environmental issues and of the challenges associated with governing them. The characterisation of environmental problems as 'wicked' (Rittel & Webber, 1973), 'messy' (Ackoff, 1974) or 'ill-structured' (Geurts & Vennix, 1989; Mason & Mitroff, 1981) has become so pervasive that it is difficult to imagine any contemporary policy issue not being labelled as such. In an attempt to pin down the core of the complexity notion, Axelrod & Cohen (1999, 7) argue that a system is complex 'when there are strong interactions among its elements, so that current events heavily influence the probabilities of many kinds of later events'. This rings true for both the cognitive or analytical and the social dimension of complexity. Clark (1986) has characterised these two dimensions by resorting to the metaphor of the earth as a garden: not only is it difficult to determine what kind of garden is possible (analytical complexity), but also what garden we actually want (social complexity). Analytical complexity relates to the lack of certainty of knowledge surrounding complex issues (Roelofs, 2000). Variables characterising the problem are interrelated and subject to multiple feedback loops, problems themselves are interrelated, and available choices for intervention, their impacts and possible side-effects are uncertain or rapidly changing. Environmental problems are also complex socially. Interests diverge between those responsible and those affected. Perceptions differ on the causes of a problem and what should be done about it, resulting in political conflicts. The differing underlying values that are at the root of these disagreements are an inherent part of this latter dimension (Hickson, 1986; Roelofs, 2000).

Even if the label of a ´complex problem´ may sometimes be too readily used, few would doubt that this characterisation applies to global environmental change issues, and particularly to climate change. To Jordan et al. (2010, 4), the latter represents a wicked issue 'par excellence' and Levin et al. (2012, 123ff) even call it 'super-wicked'. Challenges relating to time and scale loom especially large in the context of climate change. Given current emission trends and the possibility and likelihood of reaching unexpected tipping points (Stocker et al., 2013), taking action is urgent, and

humankind cannot afford to wait until all uncertainties and value disagreements have been resolved. Yet planning for climate change often goes beyond the time horizon of policy-makers or organisations, and some impacts will not be felt for a long time to come. In terms of scale, effective mitigation (reduction of greenhouse gas emissions) requires behavioural changes from actors around the globe, creating significant collective actions problems. This also implies major equity dilemmas. The adverse effects of climate change will be most dire for those in the global South that have hardly contributed to it. At the same time, most large emitters can shield themselves relatively well from climate impacts.

The need to deal with complex environmental problems has not been without consequences for the realms of policy and scientific research. For both, it has exposed the limits of traditional 'rational', unitary approaches and led to the involvement of new actor groups into core processes of knowledge production and governing. For science, the difficulty of unambiguously defining complex problems and of determining the relevant variables has led to calls for reviews of scientific processes and outcomes by an 'extended peer community' (Funtowicz & Ravetz, 1993, 753). This model of knowledge production, also referred to as 'mode two science' (Gibbons et al., 1994), emphasises the need for input from scientists from other disciplines and stakeholders to enhance the social robustness of the research being conducted.

In the realm of policy, there is now widespread recognition that, in addition to public policies (actions taken by governments in response to specific circumstances, with the overarching objective to expand the public good, see e.g. Howlett, 2011), it is governance approaches – rule-making in a variety of forms, from regulatory to network- and market-based, and by a variety of actors, public, hybrid and private – that are shaping the responses to contemporary environmental problems. And even where the government is the dominant actor, the importance of stakeholder-driven processes to inform relevant choices and decisions is increasingly recognised. These trends raise several questions as to their normative implications and practical implementation. Yet their significance cannot be overestimated. They are rooted in the notion that a good way of characterising the simultaneous processes of sense-making and social negotiation is as learning. And that, given the analytical and social complexity of environmental problems, there is a need for such learning to go above and beyond what traditional modes of learning have been able to deliver (Armitage et al., 2009; Keen et al., 2005; Parson & Clark, 1995).

The assumption that the processes necessary to effectively govern complex environmental problems can be characterised as learning is the starting point for this thesis. The emerging architecture of assessments, appraisals and engagement processes around environmental issues represents an attempt at building learning processes into environmental governance that are more structured and collaborative

and that help integrate the plurality of relevant views, norms, facts and uncertainties into decision-making. Yet while such processes have been implemented frequently over the past years, in different contexts and at different scales (Turnpenny et al., 2009), there has been little systematic evaluation of the resulting learning effects. In this thesis, we aim to make headway with closing this gap in the literature by developing a robust, empirically workable conceptualisation of learning in and for environmental governance and a set of instruments to measure it. More specifically, we examine the learning effects of the policy exercise, a sub-form of simulation games that has repeatedly been used in a global change context and that has been associated with significant learning effects in the literature.

The following sections address the need for learning in environmental governance and scholarly interest in the concept (1.2), outline different approaches to learning in environmental governance (1.3) and review difficulties with its conceptualisation and measurement (1.4). Section 1.5 presents the research questions and the approach taken in this thesis to address them. Section 1.6 contains an outline of the rest of this book.

1.2 The need for learning

The first to use the term 'learning our way out' of unsustainability was arguably Lester Milbrath. He used it as the title of his 1989 book, which was among the first to connect social learning and sustainable development (Milbrath, 1989). Since, it has become the standard recommendation for dealing with complex environmental problems; also the International Panel on Climate Change (IPCC) calls 'the institutionalization of learning among actors […] an important aspect of success' in developing more robust international climate policy (Stavins et al., 2014, 1010). Learning is needed regarding both the analytical and the social dimension of complexity: it is as much about gaining insights into drivers, dynamics and impacts of environmental problems and possible ways of addressing them as it is about understanding possible opposition to certain solutions, its implications and feasible alternatives. Learning will have to happen across multiple levels of governance and across the society as a whole, extending from the individual to local communities, corporations and public actors. Yet doing so successfully also requires more thorough study of learning processes themselves – how they can be triggered, what their dynamics are, and how their results can be consolidated (Blackmore, 2007; Lee, 1999; Leeuwis & Pyburn, 2002).

Scholars in the policy sciences have been interested in the notion of learning in policy (or 'policy learning' or 'policy-oriented learning', as some prefer to call it) for some time. It is part of the larger debate on the role of power, interests and ideas in policy-making. Despite Max Weber's famous quote emphasising the role of ideas, which

'like switchmen, determined the tracks along which action has been pushed by the dynamic of interest' (Weber, 2007 [1946], 280), early public policy theorists explained policy change through a primarily conflict-based lens – as a power struggle between different interests in the policy arena (Easton, 1953, 1965; cf. Freeman, 2006; Grin & Loeber, 2006; Parsons, 1995).

Only from the 1960s onwards theorists increasingly acknowledged that changes in policy may also be rooted in shifts at the ideational level, and found the concept of learning to be of added value in explaining these processes. Those in power may incorporate new ideas or lessons learned from previous experience or other countries into their day-to-day activities (Etheredge, 1981; Heclo, 1974; Schön, 1973), or their paradigms and values may change as a result of broader shifts in society (Hall, 1993). At an empirical level, Freeman (2006) attributes the increased interest in learning among policy-makers in the 1960s and 70s to two factors. First, the rapid social, political, economic and technological changes taking place led to a rising sense of uncertainty about appropriate governmental interventions. Second, there was a growing awareness that policy problems and possible solutions might be similar across countries, and that looking beyond national borders to compare experiences and learn from each other might prove beneficial. The following section delves in more detail into different approaches to learning in environmental governance.

1.3 Approaches to learning in environmental governance

At the most basic level, learning in environmental policy occurs ad hoc, as policy-makers or other actors react to outside events and stimuli, and adjust their actions on this basis (Kingdon, 1984; Lindblom, 1959). This learning can be based on direct experience or can build on observations from other policy domains, other countries or different points in time (May, 1992). While beneficial in many instances, it may not be sufficient for dealing with the multiple facets of complex environmental problems as outlined above. Nonetheless, from an analytical viewpoint, it is worthwhile examining this ´default´ mode of learning as it unfolds in practice, for instance by means of process tracing approaches, with a view to identifying driving forces and conditioning factors (Busenberg, 2001; Nilsson, 2005).

The recognition that complex environmental problems may require specific methods to trigger learning is at the basis of the literature on adaptive management, and, more recently, adaptive co-management. Proponents of adaptive management emphasise the benefits of 'learning by doing' (Holling, 1978; Lee, 1993, 1999). They acknowledge that policy-makers know little in advance about the effects of their interventions and thus emphasise the importance of incorporating the lessons learned into the next iteration of planning. They call for treating these interventions as large-scale experiments, with an emphasis on monitoring and evaluation, and also

on policy termination if necessary (Walters & Holling, 1990). Adaptive co-management adds to this a focus on the role of communities and stakeholders in managing natural resources and to the importance of horizontal and vertical linkages in policy-making (Armitage et al., 2008; Armitage et al., 2009). Adaptive co-management has been heralded by its proponents as an appropriate strategy for dealing with the uncertainties and complexities inherent in environmental change. Yet authors have also highlighted problems associated with its implementation (Gregory et al., 2006). For one, it may be more suited to smaller-scale interventions that lend themselves to centralised monitoring. Secondly, while entirely in the spirit of adaptive management, the 'double-edged sword of 'successful failures'[1] has served as an institutional, political, and emotional barrier to the implementation of adaptive management' (National Research Council, 2009, 76). Finally, the long-time horizons involved in establishing long-lived infrastructures, for instance for climate mitigation, limit the applicability of these approaches. An active scholarly community continues to research the constraints and opportunities embodied in in adaptive co-management approaches, seeking to gradually transcend the traditionally dominant case study approach in favour of broader lessons learned (see e.g. Plummer et al., 2012).

Ex-post evaluation of programmes or policies – also practiced in adaptive management – is a widely applied group of approaches aimed at stimulating learning in environmental governance. In a meta-analysis of evaluation studies of climate policy, we observed how the increase in policy activity since the adoption of the Kyoto Protocol has been accompanied by a corresponding rise in evaluation activity in six European countries and at the EU level (Haug et al., 2010; Huitema et al., 2011). Yet when attempting to distil the broader lessons emanating from these evaluations for more effective future climate policy, we noted that the substantive evidence base that these evaluations represented was surprisingly weak (Haug et al., 2010). The large majority of studies were neither reflexive (i.e. questioning official policy goals) nor participatory in character, and very few attempted to ascertain the ultimate effectiveness of a policy in terms of emission reductions, arguably a key goal for climate mitigation measures (Haug et al., 2010; Huitema et al., 2011). Current climate policy evaluation practice in Europe thus falls short on three important counts – degree of participation, reflexivity, and focus on outcome effectiveness –, thereby severely limiting possibilities for policy learning. The study did not even examine the uptake of evaluation outcomes in policy-making, the last but crucial step to ensure that the learning embodied in the evaluation outcomes eventually translates into policy practice. This example illustrates that there is scope for more and better evaluation of the measures designed to address complex problems, as well as for studying how evaluations are carried out and how – if at all – they feed into policy-

[1] [experiments that are deliberately terminated as they failed to meet the expectations formulated at the outset, C.H.]

making. Even once improved, however, the potential usefulness of ex post evaluation is likely to be limited especially for climate change adaptation. Not only is adaptation merely in its infancy as a policy field (Massey & Huitema, 2012), judging the effectiveness of measures is also complicated by the time lag with which many climate impacts are likely to be felt. And in some instances, the insights gained from ex post evaluation might simply come too late.

This is where a fourth group of approaches – forward-looking ex ante assessment or appraisal – comes in. Owens et al. (2004, 1944) define them as 'that family of ex ante techniques and procedures ... that seek to inform decision makers by predicting and evaluating the consequences of various activities according to certain conventions.' Like evaluation, ex ante assessment can occur in different ways. It may be conducted through institutionalised systems directly linked to political decision-making processes, like environmental impact assessment (EIA) or strategic environmental assessment (SEA) (see e.g. Turnpenny et al., 2009) or in independent research efforts with the aim of producing policy-relevant insights that are not commissioned by an official body. As for ex post evaluation, the learning benefits from an appraisal depend not only on its design and quality. They also hinge on the surrounding politics, which may prevent the results of an otherwise successful process from being taking up in policy-making (Turnpenny et al., 2009). Method-wise, there is a wide spectrum from purely desk-based, often quantitative studies to more participatory and deliberative approaches (National Research Council, 2009). The US National Research Council argued that in the context of climate change, 'collaborative, broadly based, integrated, and iterative analytic-deliberative processes' provide the method of choice for organising scientific analysis to serve public decision making' (National Research Council, 2009, 79). Such an approach to appraisal resembles the 'participatory integrated assessment' (PIA) processes that have been implemented since the late 1990s (see Chapter 2).

In this thesis, we examine the potential for policy-oriented learning from one particular type of ex ante assessment, the policy exercise. Policy exercises, a form of simulation-games, aim to produce policy-relevant insights by placing policy-makers and experts in a future scenario context, where they adopt the roles of key policy actors and negotiate and take decisions in a structured process. There are great claims in the literature regarding the learning potential of such approaches to policy development in complex, multi-actor settings (Duke & Geurts, 2004; Parson, 1997; Underwood & Duke, 1987). The pressure on participants to effectively share information and act under time pressure as well as the safe environment that a game constitutes and that allows for experimenting with different policy options, are often cited as important aspects in this regard (e.g. Parson, 1996b). However, as for other forms of appraisal, such claims are hardly supported by systematic empirical evaluations of the learning effects that may be generated (see next section). We will

come back to the policy exercise method, its premises, background and history in Chapter 2.

1.4 Learning for environmental governance: conceptual ambiguities and the need for a better evidence base

Learning in and for environmental governance is not a new topic of research. There has been a wealth of analyses examining learning processes and outcomes across different contexts and social units of analysis. Studies draw on learning theories and typologies from a range of disciplines. Yet also after more than twenty years, the literature continues to struggle with major conceptual and empirical challenges. The problems that authors encounter vary with the focus of the studies, ranging from broad accounts of learning in a policy field over several years to case studies tracing learning in a community or by individuals as a result of an appraisal process or other events. In a review of learning in foreign policy from 1994, Levy (1994) concluded that the 'concept of learning is difficult to define, isolate, measure, and apply empirically.' This still rings true for the environmental field today.

At the core of these difficulties is the lack of a widely accepted, empirically workable definition of learning in environmental governance. Looking at the concept of social learning, Reed et al. (2010, no page number) deplore that some definitions employed in the literature 'are so broad they could encompass almost any social process.' This vagueness renders focused empirical analysis and cross-case comparison exceedingly difficult. Moreover, learning processes and learning outcomes are often conflated, or one is implicitly privileged or exclusively treated at the expense of the other. In studies focusing on smaller units of analysis, 'social learning as a concept is frequently confused with the conditions or methods necessary to facilitate social learning, such as stakeholder participation' (Reed et al., 2010, no page number). Indeed, participation may be a precondition or helpful mediating factor for social learning to occur, but this thesis argues that it is not synonymous with it. There are also many cases in which learning is equated with its outcomes, such as observed behavioural shifts or improved environmental management for smaller-scale studies, or policy change for larger scale ones. Yet again, the occurrence of such changes hardly constitutes conclusive evidence for learning as such, as they may also be due to other factors (May, 1992). We believe that a narrower definition of learning would help clarifying the relationship between independent and dependent variables.

Another potentially problematic feature of many popular learning typologies is the more or less explicit hierarchy between different types of learning. Typologies tend to treat normative aspects of learning as superior to cognitive ones. This generalisation however may not always hold true and risks complicating empirical analysis of learning phenomena (see Chapter 2 for a more in-depth discussion). Finally, some

studies only consider as learning those lessons that the authors deem desirable. This normative basis is particularly problematic when the underlying assumptions on which an assessment is based are not clarified ex ante (Diduck, 2010; Levy, 1994).

Our own conceptualisation of policy-oriented learning in environmental governance aims to remedy the above-mentioned challenges. Regarding the normative bias inherent in many learning definitions, this thesis argues that studies examining learning in environmental governance should primarily look out for a 'change in understanding' among the subjects concerned (see Chapter 2).[2] It is the change that is important to us here, not its direction. An assessment of the potential implications of the learning achieved can always follow as an – analytically distinct – next step. Furthermore, we move away from the ´layered' nature of learning embraced by other typologies. Instead, we propose a three-pronged typology that distinguishes between cognitive, normative, and relational learning. Cognitive learning refers to the acquisition of new or the better structuring of existing knowledge, normative learning implies a change in views, values or norms, and relational learning a better understanding of other's mind sets and enhanced trust and ability to cooperate. We believe that this typology constitutes an improvement over existing approaches used in an environmental governance context. It captures its key dimensions and treats these equally rather than privileging one over the other (while admitting for the possibility of the interrelatedness of the different types). Finally, it is intuitive and sufficiently precise to identify phenomena that are empirically observable.

The problems with the study of policy-oriented learning in environmental governance do not end with issues in the conceptual realm. Challenges on the empirical side are just as pronounced, if not more. With a few exceptions (Baird et al., 2014; Deyle & Schively Slotterback, 2009; Garmendia & Stagl, 2010; Huitema et al., 2010), studies that systematically evaluate and measure learning outcomes in environmental governance are surprisingly scarce. This is to do with the vagueness of the learning notion, but certainly also with the way in which evaluations in this area are conducted. As Rodela et al. (2012) conclude from a systematic review of 54 papers studying social learning, there is a tendency towards loosely reflective ex post studies where authors use ´hindsight´ to analyse the – exploited or missed – learning opportunities offered by a participatory process or other project. They attribute this to the relatively recent interest in the notion of social learning, which prompted authors to retroactively apply a ´learning lens´ to projects and data collection efforts originally designed with other objectives (and hence research questions) in mind. There is certainly a place for these studies as the reflections contained in them provide useful lessons for the design of future efforts in the same field. But to really make sense of the phenomenon of policy-oriented learning in environmental

[2] This does not resolve the dilemma of how to conceptualise the collective property of learning in group processes as opposed to individual learning, however (see Chapter 2).

governance, of what triggers and what impedes it, there is a critical need to move beyond single-case studies relying exclusively on qualitative, ex-post analysis, towards controlled (quasi-) experiments (see Chapter 3). This is the challenge that this thesis takes on.

1.5 Research approach and questions

We have thus far established that many environmental problems, especially in the context of global environmental change, are both analytically and socially complex. Addressing this complexity through adequate policies requires a significant extent of learning, by policy-makers, scientists, and eventually the public at large. Some of this learning happens unplanned and ad hoc. Yet this will not be sufficient for issues like climate change that pose unprecedented challenges in terms of scale and time. This is where approaches specifically geared towards learning, like ex post evaluation and ex ante appraisal, as well as approaches advocated in the context of adaptive (co-) management may have value. Yet they each present their own set of challenges, whether it is the type of lessons that can be drawn from them or their incorporation into policy-making. A better understanding of these procedures, their opportunities and limitations thus represents an important stepping stone towards more robust and ultimately also more legitimate policies to address climate change.

In short, the goal of the present thesis is to contribute to a better understanding of policy-oriented learning in and for environmental governance, at both the conceptual and the measurement level. Using two policy exercises as an empirical case to test out our conceptualisation of learning and the instruments developed to evaluate it, we also aim to add knowledge on the opportunities and limitations of one specific form of appraisal that is frequently connected with significant learning effects in the literature. This brings us to the following three research questions: *how can we better conceptualise policy-oriented learning in an environmental governance context? How can this concept be operationalised and measured empirically? And what prospects do policy games, more specifically policy exercises, offer for learning in this regard?*

Turning to our empirical case, we designed and ran two policy exercises on subjects in the realm of international climate policy. The first one was about challenges related distributing emission reduction commitments among member states in the European Union, a process commonly known as 'burden-sharing' or 'effort sharing' (see e.g. Haug & Jordan, 2010; Ringius, 1999). The second exercise examined options to set up a global fund for REDD (Reducing Emissions from Deforestation and Forest Degradation) under the international climate regime (see e.g. Haug & Gupta, 2013; Streck et al., 2009). Such a fund would offer financial compensation to developing countries that succeed in protecting their forests, thereby avoiding emissions from lost forest carbon stocks. We ran both policy exercises first with groups of

international experts and policy-makers, and then with students in the Environment and Resource Management Masters programme at the Vrije Universiteit Amsterdam. In order to measure the cognitive, normative and relational learning effects from the exercises, we applied a variety of methods, including surveys, interviews and participant observation as well as pre- and post-concept maps drawn by participants.[3] The multiple-case approach (measuring learning in four different participant groups), the combination of evaluation methods and the longitudinal nature of the assessment, with measurements taken at multiple points in time, clearly go beyond the current state of the art in assessing learning from environmental appraisals (Blackstock et al., 2007; Garmendia & Stagl, 2010).

1.6 The rest of this thesis

In the quest for a better conceptualisation of learning in environmental governance and how to measure it, this thesis proceeds as follows.

Chapter 2 provides the theoretical underpinnings of this research through a brief survey of the diverse, multidisciplinary literature on learning and a thorough introduction to simulation gaming and policy exercises. It presents the diversity of views as to who learns in public policy and environmental management, what is learned and to what effect. The question what is learned leads us to distil what we consider the three key dimensions of learning – cognitive, normative, and relational – that form the basis for our own conceptualisation. We then describe background and rationale of simulation gaming and the policy exercise method, the learning effects of which we investigate in the empirical part of this thesis, and its possible functions for policy development.

Chapter 3 highlights the methodological choices made in this thesis. It addresses case selection, the considerations underlying the evaluation framework used to assess the learning effects from the policy exercises and its elements, and the overall research design. It also discusses limitations in terms of research design and data availability.

Chapter 4 describes in more detail the two policy exercises that we organised. Brief reflections on good practice in the design of simulation games and on the specific approach adopted in this regard are followed by an introduction to the policy context of the two exercises. The major part of this chapter is then dedicated to presenting the design, runs and outcomes of each exercise.

[3] Concept maps graphically depict a person´s understanding of a topic at a given point in time (Novak & Gowin, 1984; Novak, 1998). Comparing the maps drawn by participants prior to and after the policy exercise allowed us to trace the conceptual shifts and hence cognitive learning as a consequence of the intervention (see Chapters 3 and 5).

Chapters 5, 6, and 7 form the empirical core of this thesis as they analyse the learning effects that we observed from the policy exercises. Chapter 5 focuses on participants' cognitive learning, drawing on data from the pre- and post- concept maps as well as self-reported data from surveys and interviews. Chapter 6 examines the extent to which participants learned in normative terms, comparing and cross-checking self-reported and recorded data from surveys and interviews. Chapter 7 finally addresses the relational dimension, focusing primarily on the question to what extent participants gained a better understanding of others' mind sets, including of the roles they adopted during the policy exercise.

The concluding Chapter 8 draws the findings of this thesis together. It first synthesises the learning effects found, assesses factors that may have stimulated or hindered participants' learning, and discusses some lessons learned for the organising team of the policy exercises. This is followed by a broader discussion on the role and potential of policy exercises as an appraisal method for addressing complex environmental problems. The chapter then reflect on the usefulness and limitations of the typology of learning developed in this thesis and the associated evaluation framework. Finally, we outline a research agenda, before closing with some reflections on the state of international climate governance.

2. Learning for policy and the role of policy games in policy appraisal

This chapter embeds the present research into its wider context and explain the most relevant concepts and the theory behind it. The first part introduces the notion of and rationale for learning in public policy in the sphere of environmental governance and examines how learning is conceptualised in the diverse and multidisciplinary literature that exists on this topic. Building upon these literatures, we then present our own alternative typology of policy-oriented learning in environmental governance, which is broadly applicable and useful for testing across a wide range of interventions.

The chapter goes on to introduce simulation-gaming, and a sub-form of it, the policy exercise, the learning effects of which we measure in the empirical part of this thesis. We trace the historical evolution of simulation-gaming, describe the characteristics of the policy exercise method and previous applications in the domain of climate policy, and end with a summary of the claims made about the learning benefits from policy games.

2.1 Learning for policy

As Chapter 1 argued, learning is increasingly becoming a normative goal in the face of complex environmental problems. A rich and multifaceted literature exists on learning in a social context, stemming from disciplines as diverse as social psychology, education studies, organisation and management, policy sciences, international relations and environmental and natural resource management. Among these, the bodies of work on policy learning and on social learning are most relevant to informing this thesis.

Interest in the notion of *policy learning* first developed in the 1960s when authors sought to supplement and counter existing theories focusing primarily on conflict as a catalyst for policy change (Bennett & Howlett, 1992). As Hugh Heclo (1974, 305-306) famously observed,

> *'Politics finds its sources not only in power but also in uncertainty — men collectively wondering what to do [...] Governments not only 'power' [...] they also puzzle. Policy making is a form of collective puzzlement on society's behalf; it entails both deciding and knowing [...] Much political interaction has constituted a process of social learning expressed through policy.'*

Core notions of the policy learning literature will be more closely examined in the following sections. The concept of *social learning* in turn finds its early roots with Bandura (1977), who defined it as the continuous and iterative interaction of an individual with the outside world. This initial focus was soon expanded to become part of a broader paradigm that emphasises the importance of joint, interactive learning of actors as a key element to an interactive, participatory approach to decision making and problem solving, often in the context of natural resources management (Leeuwis & Pyburn, 2002; Muro & Jeffrey, 2012; Pahl-Wostl et al., 2007; Schusler et al., 2003). Social learning literature has strongly informed our notion of relational learning, which is discussed further in Section 2.1.3.

Interest in the concept of learning is thus widely shared; yet authors look at largely distinct phenomena, and at different levels, when analysing learning. The observation by Parson and Clark (1995, 429) that the literature shares neither a common disciplinary background nor a theoretical perspective rings as true today as it did almost twenty years ago. A review of the key themes and perspectives present in the literature on learning in an environmental governance context thus constitutes a key stepping stone before delving into developing our own conceptualisation (see Section 2.1.4). In their review of developments in the policy learning literature until the early 1990s, Bennett and Howlett (1992) use three axes to organise the field: who learns; what is learned; and to what effect. We emulate this effective structure in the following sections. Before turning to the question what is learned, which will remain a central theme throughout this thesis, we first discuss the other two questions – 'who learns' and 'to what effect'.

2.1.1 Who learns? The issue of the social unit of analysis

Policy-oriented learning happens at different scales; Diduck (2010), in his effort at 'untangling the learning dimension of adaptive capacity', distinguishes between five different conceptions of learning, and related units of analysis: Individual learning, action group learning, organisational learning, network learning, and societal learning.

At the largest scale, societal learning, accounts focus on 'democratic processes by which core societal institutions are modified in response to social and environmental change ' (Diduck, 2010). Authors differ, however, in whom exactly they identify to be the agents of change, or who 'the society' in question is composed of. At one end of the spectrum, there are authors who focus on learning by governmental organisations and their leaders (e.g. Etheredge, 1981). Others (e.g. Hall, 1993) observe learning processes as complex interactions between state officials and societal forces. According to a last group, most prominently represented by Sabatier, learning occurs in a policy subsystem within and across different advocacy coalitions (Sabatier, 1987, 1988, 1998). Medium-scale analyses examine learning in networks and organisations, including the notion of the 'learning organization' (Senge, 1990)

and the features and approaches that facilitate or impede it. Finally, a substantial body of work focuses on smaller social units of analysis, such as groups or individuals, who learn in the context of participatory processes or workshops.

The unit of analysis also touches on the question concerning the relationship between individual and collective learning. This is a dilemma which many authors avoid to make explicit or touch upon only superficially. People learn individually. Yet at the same time, many authors, especially in the social and organisational learning fields, consider learning an emergent property from social interaction and use it as a convenient shorthand for a collective where learning is taking place. Especially in organisational science, authors place emphasis on this collective element. According to them, most learning in organisations takes place through the incremental adaptation of routines (Cyert & March, 1963) that are engrained in the collective and are 'independent of the individual actors who execute them' (Levitt & March, 1988, 320).

For Eiser (1986), choosing between collective and individual learning comes down to a philosophical difference in approach: some authors regard the social context as merely an external stimulus to learning by individuals, whereas others see it as an arena for individuals to learn together and develop shared meaning. In the context of organisational learning, Argyris and Schön (1980) describe organisational learning as a metaphor: 'organizational learning is not the same thing as individual learning […] Organizations do not literally remember, think, or learn […] There is no organizational learning without individual learning' (Argyris and Schön, 1980, 9-11, 20). Hedberg (1981, 3) concurs: 'it is individuals who act and who learn from acting; organizations are the stages where acting takes place.' This leads Levy (Levy, 1994, 291) to argue for an analysis of learning at the level of the individual, cautioning that the 'the reification of learning to the collective level [...] is not analytically viable.'

While acknowledging the potential 'emergent property' nature of some learning events, especially when embedded in an existing social network, this thesis nonetheless opts for a focus on learning at the individual level. This is partly because any potential collective learning rests upon individual learning as a pre-condition, even according to the proponents of collective learning approaches. Moreover, in view of the empirics of this research project, which brings together individuals for a one-off event, a focus on the collective did not appear adequate.

2.1.2 Learning to what effect?

Another guidepost that Bennett and Howlett (1992) use is the question 'to what effect' learning is taking place. In the context of policy learning, the occurrence of learning is often linked to concrete evidence for policy or institutional change (e.g. Bennett & Howlett, 1992; Busenberg, 2001; Jarosz & Nye, 1993). For some authors

this is a defining element: 'learning occurs when understanding, insight and explanations are connected with action' (Argyris, 2003, 1179). However, directly equating learning with policy change entails a risk of empirical bias. If the occurrence of policy change determines the cases that are examined, one risks missing out others where learning happened, but change was inhibited by other factors – which would be very interesting to investigate in their own right. Furthermore, in some instances, learning may also serve to validate existing policies when the knowledge base for a certain course of action has improved (Huitema et al., 2007; Levy, 1994) – or it may simply not translate in a change that can easily be observed from the outside (Mathews, 1994).

Requiring learning to be linked to observable policy change gives rise to other pitfalls. For one, making the connection between learning, for instance through a participatory process, and policy change that can be clearly traced back to the 'learning event' is difficult. Even where such events are clearly linked to a decision-making context, their impact on policy-making is far from assured. Thus, Huitema et al. (2010) argue that these processes are better evaluated against the goals that they were set to achieve (for instance raising awareness, or improving the quality of decision-making) than against policy change as the sole and most important yardstick. The social learning literature is divided on the question whether resulting action – not necessarily equating policy change – is a necessary condition for ascertaining social learning in the first place. Some authors like Pahl-Wostl (2006) and Blackmore (2007) insist that social learning needs to transcend the cognitive realm and result in joint practices and collective action. To us, this confuses the process with the desired end result. Learning occurs as individuals' understanding of the situation, their belief systems and intentions to act change as a consequence of their experience; whether this leads to changed behaviour is a distinct question as other intervening variables independent from the learning process may prevent it (Muro & Jeffrey, 2008; Reed et al., 2010). In the following, we will therefore conceptualise learning as a change in understanding.

The definition of what counts as such a 'change in understanding' also begs the question of the 'directionality' of learning. Parson and Clark (1995, 456) put it this way:

> *'What counts as learning? Any change in the phenomena studied, or only certain changes? If only certain changes, then what criteria distinguish learning change from nonlearning change? Short of calling any change learning, one might say that only change in response to identifiable stimuli or information is learning; or more restrictively still, that the change must follow rationally (defined somehow) from the stimulus.'*

For many authors writing about social learning from an adaptive (co)management perspective or in an action research context, social learning per definition contains a

normative prescription. They implicitly or explicitly equate learning with 'learning towards sustainability', in its broadest or in a narrower sense (e.g. Garmendia & Stagl, 2010; Tábara & Pahl-Wostl, 2007). This line of reasoning has its equivalent in the policy learning literature, with for instance Stein arguing that an 'evaluation of the structure and content of cognitive change' is '[i]nescapably built into the concept of political learning' (Stein, 1994, 171). Argyris and Schön (1980) equally invoke the inherently normative character of theories of effective learning.

These authors apply what could be termed an 'accuracy criterion' (Levy, 1994, 294) to the concept of learning: only learning that is judged as empirically correct or desirable in normative terms counts as 'real' learning (ibid.). The upside to such an approach may be that it helps sharpen the analytical focus as to what empirical phenomena to look out for. Yet it entails inevitable biases as observers evaluate outcomes on the basis of their own frames and normative assumptions (Levy, 1994). At a minimum, this calls for the criteria and underlying assumptions to be made explicit in the studies using such an approach. This, however, is rarely the case in practice. The lack of widely accepted standards to evaluate learning based on criteria such as accuracy, efficiency or sustainability leads Diduck (2010, 206) to warn of a 'real danger of analysts using the term learning in so self-serving a manner that it becomes synonymous with the accrual of outcomes that the analyst deems correct.' For these reasons, we will consider any change in understanding as learning and not assume any particular direction of change in this study (Nilsson & Swartling, 2009).

Having briefly sketched out the various angles and perspectives under which Bennett's and Howlett's questions on the learning agent and the effects of learning have been treated in the literature, we will now turn to their third question - the object of learning. What exactly is it that authors are talking about when they are analysing policy learning and social learning?

2.1.3 What is learned - the object of learning

In the policy learning, but also in the social learning literature, the object of learning has mostly been discussed in terms of different levels or types of learning. Table 1 presents an overview of the conceptualisations of a number of authors from the policy, organisational and social learning literatures. Other scholars use the same concepts or terms with slight adjustments.

Implicit in all typologies is a hierarchy from the first – more technical, less reflexive – level of learning to higher-order forms. Furthermore, all frameworks contain a more cognitive and a more normative, critical learning dimension, the latter often being associated with the higher learning levels. Whereas most authors distinguish between two levels, some frameworks also add a third. In some cases (e.g. Argyris & Schön, 1980), that third level is associated with a meta-dimension ('learning how to

learn'),[4] while in others it contains an action-based orientation which effectively connects this highest learning level with (deep) policy change (Armitage et al., 2008; Pahl-Wostl, 2009).

Relevant studies	**Learning concept employed**	**Levels of learning identified**
Lindblom (1968)	Policy learning	*Instrumental learning*: new insights into policy options in the case of a given policy problem and a given solution context *Political learning:* new insights not only into solutions, but also the problem itself and the context in which decisions take place
Argyris and Schön (1980, 1996)	Social learning	*Single-loop learning:* discovering and correcting errors without changing underlying norms, policies and objectives *Double-loop learning:* discovering and correcting errors in a way that does involve the modification of underlying norms, policies and objectives *Deutero-learning*: 'learning how to learn' (see footnote on previous page).
Senge (1990)	Organizational learning	*Adaptive learning*: 'survival learning', focuses on solving problems in the present without examining the appropriateness of current learning behaviours *Generative learning:* learning that requires systems thinking, emphasising continuous experimentation and feedback on the way how problems are defined and solved.
Hall (1993)	Policy learning	*First-order learning*: learning about the settings of policy instruments *Second-order learning*: learning about techniques or policy instruments used to attain certain policy goals *Third-order learning*: new insights into overall goals of policy, amounting to a paradigm change
Mezirow (1995)	Transformative learning	*Instrumental learning:* the acquisition of skills and knowledge (the 'how' and the 'what') *Communicative learning:* reflection on the assumptions imbedded in the social discourse

[4] Argyris' and Schön's definition of deutero learning contains some conceptual ambiguities. Sometimes, it is described as a meta-dimension to both single-loop and double-loop learning (e.g. Argyris, 2003), whereas elsewhere (e.g. Argyris & Schön, 1996), it appears to be more of a form of double-loop learning. Similarly, it remains unclear if the term refers to the process of reflection and inquiry, or the structural conditions that make such an inquiry possible (cf. Visser, 2007).

		('learning what people mean') *Transformative learning: '*The process of using a prior interpretation to construe a new or revised interpretation of the meaning of one's experience in order to guide future action' (Mezirow, 1996, 162)
Glasbergen (1996)	Social learning	*Technical learning:* search for new policy instruments in the context of fixed policy objectives *Conceptual learning*: Redefining policy goals and adjusting problem definitions and strategies
Armitage et al. (2008), building on King and Jiggins (2002) and Keen et al. (2005)	Social learning	*Single-loop learning:* Identification of alternative strategies and actions to resolve specific problems and improve certain outcomes *Double-loop learning:* Challenging existing worldviews and underlying values *Triple-loop learning:* amending underlying governance norms and protocols

Table 1. Levels of learning across the social, policy and organizational learning literatures.

Exactly what kind of learning is located at what level differs to some extent between the frameworks (cf. Armitage et al., 2008, 89). While for Hall (1993) for instance, consideration of alternative policy instruments already implies second-order learning, others (e.g. Glasbergen, 1996) subsume this under the first level of learning. Similarly, in some typologies, the first level is exclusively associated with cognitive learning; in others, cognitive and normative elements are mixed up across the framework. Last but not least, sometimes, the types are simply used as different categories by which to characterise learning processes, whereas other authors (e.g. Glasbergen, 1996; Pahl-Wostl, 2009) see them as both evolutionary (one kind of learning leads to the other) and cumulative (one builds on the experience of the former).

Even though a more or less implicit hierarchy moving from lower-order to higher-order learning types is ubiquitous in the literature, such an understanding of learning poses a number of difficulties. First, it seems doubtful that the normative, reflective type of learning is inherently more valuable than more cognitive types (as in the definition of Armitage et al., 2008, in the table above). Cognitive learning on its own may well generate fundamental effects (cf. Owens, 2004): for instance, a different perception of the problem (a cognitive process) may not result in a normative shift, but enable a virtuous cycle of yet more cognitive learning. At the same time, normative learning may not always be beneficial to problem solving. Taking an example of international climate policy, giving up on the hope that big emitters like China eventually play their part in greenhouse gas mitigation may reduce motivation to take action at home. This normative shift consequently slims rather than increases the chance of successfully tackling climate change. Also the interrelation between cognitive and normative learning is more complex than the one-way connection of

cognitive learning giving rise to normative learning assumed by some authors. In the reverse, normative learning may also open opportunities for cognitive progress. Again using the example of international climate policy, coming to accept the right of developing countries to grow their emissions might lead to devising a more positive, inducement-based approach. Finally, the implicit preference for normative learning risks inducing a bias in empirical research, leading authors to focus on seeking out evidence for normative shifts, while potentially neglecting cognitive changes which may be just as important and desirable. In sum, there is a need for more neutral wording when describing types of learning. For this reason, the learning typology developed in this thesis treats normative and cognitive learning on an equal footing, without privileging one over the other (see Section 2.1.4).

Moreover, it is interesting to note that current conceptualisations – even the ones that focus on social learning – do not explicitly distinguish an important aspect of learning, which is relational. Relational learning, which focuses on the relationship between individuals in a learning situation, is mentioned in the literature, yet it is often conflated with other forms or levels of learning. Hisschemöller et al. (2001) and Pahl-Wostl (2009) for instance assume that 'higher' levels of learning combine at once changes in normative frameworks and improved relations or better understandings between stakeholders and participants. Conceptually, this does not seem very logical. Relational forms of learning can occur independently from other forms of learning (one can despise a teacher and continue to do so for a long time, but learn a lot from him or her in cognitive terms). Moreover, developing a better awareness and understanding of how another stakeholder thinks might enhance the respect for that stakeholder, but does not necessarily result in changes in one's own normative framework.

This leads us to propose a three-pronged typology of learning effects, differentiating between cognitive, normative and relational learning (see Haug et al., 2011; Huitema, et al., 2010; compare Webler et al., 1995). In the following three sections, we will explore the elements and outcomes commonly associated with each of these three dimensions in greater detail, before distilling from this our own definition of the learning types.

2.1.3.1 The cognitive dimension of learning

In any social context, the transmission of information between individuals is a core component of the cognitive dimension of learning (Crona & Parker, 2012; Newig & Pahl-Wostl, 2010; Reed et al., 2010). It occurs through communication and exchange between actors, which is why diversity of knowledge and expertise is often considered key to the success of appraisal processes. In an investigation into the role of knowledge for achieving environmentally friendly behaviour, Kaiser and Fuhrer (2003) distinguish between three types of relevant knowledge: (1) declarative knowledge, which includes insights into how the system works and into the state of

the problem (cf. Schahn, 1996); (2) procedural knowledge, which relates to behavioural options and to learning how to reach a particular goal; (3) and effectiveness knowledge, which refers to knowing about the relative effectiveness of different strategies to reach that same goal. Levy draws the lines somewhat differently: he differentiates between (1) causal learning – insights into cause-effect relationships and 'the optimal strategies under various conditions' (Levy, 1994, 285) and (2) diagnostic learning, 'changes in beliefs about the definition of the situation or the preferences, intentions, or relative capabilities of others' (ibid.). According to Levy (1994), any cognitive learning can be probabilistic as well as deterministic – it does not necessarily involve absolute changes in beliefs or understanding, but also changes in the degree of confidence about the knowledge that one has.

Levy's reference to cause-effect relationships and Kaiser and Fuhrer's 'insights into the state of the problem' point to a second key process in the domain of cognitive learning: the act of defining and structuring the problem. Problem structuring is a process that pertains both to individuals and collectives confronted with the need for decision-making under complexity. The process can be conceived as two counter-acting movements, 'aspect differentiation' and 'aspect integration' (Joldersma & Roelofs, 2004, 701f).[5] Aspect differentiation refers to the ability to perceive a problem in its wider context as it relates to other issues and processes. The necessity of a broad view of the problem and of its embedding into the wider system constitutes a fundamental premise of systems analysis (Geurts & Vennix, 1989). At the same time, aspect differentiation also evokes the need to take into account the variety of views and perceptions that different actors have on a particular issue, which may have significant implications for the feasibility and effectiveness for problem-solving options (Roelofs, 2000; van de Kerkhof, 2004).

Both in individuals and collectives tasked with decision-making, at some point in the problem structuring process, aspect differentiation is likely to give way to aspect integration (cf. Roelofs, 2000). Connecting loose pieces of information and identifying causal relationships provides a sense of order and transparency and enables a clearer articulation of the problem and the steps to address it (Mason & Mitroff, 1981; Rohrbaugh & Eden, 1990). This combination and integration of various aspects into a more or less coherent whole constitutes a vital process of sense-making at the level of the individual; at the level of the collective, it helps with reducing complexity so as

[5] This conceptualisation also corresponds to the notion of government learning developed by Etheredge and Short (1983). According to them, learning by state officials manifests itself in three ways: first, a greater ability for differentiation; second, a greater ability for organization and hierarchical integration. The third indicator 'increased capacity for reflective thought, perspective on the form and nature of the contents of thought, and on the choice of structuring principles' adds a reflexive meta-perspective to the first two (Etheredge and Short, 1983, 42).

to create 'processable problems' that can be assigned to specific parts of the bureaucracy (Williams & Matheny, 1995).

Greater (acceptance of) cognitive complexity and improved ability for systems thinking are often considered important goals for, or attributes of, cognitive learning. For Tetlock, one of the proponents for a cognitive structural approach to learning in foreign policy, 'learning involves change in the cognitive structure of one's image of the international environment […] in the direction of greater complexity […]' (Tetlock, 1991, 22). Yet the conclusion that increasing cognitive complexity necessarily equates learning may not always hold true. For one, a more complex cognitive structure may not actually result in a more accurate understanding of the world or in enhanced effectiveness at reaching one's goals. There may even be a risk of paralysis if aspect differentiation is not followed by integration and sense-making (Levy, 1994, 294f). Therefore, rather than increased cognitive complexity per se, the notion of more holistic or systems thinking may be a more appropriate objective or indicator for cognitive learning. Systems thinking, which implies seeking to understand where there are possible points of intervention in the system and what might be their repercussions on other part of it, is frequently evoked in the literature on social learning in a natural resource management context and is considered key to understanding complex management problems (Daniels & Walker, 1996; Kay & Schneider, 1994; Maarleveld & Dangbegnon, 2002). Garmendia and Stagl (2010) made an improved ability for holistic thinking an explicit indicator in their framework to assess social learning from participatory appraisals. Webler et al. (1995) see it as part of their criterion of 'cognitive enhancement' that such processes are supposed to stimulate among participants.

Summarising the discussion above, we can conclude that cognitive learning encompasses either the assimilation of new knowledge and facts or the improved structuring of existing knowledge, which may result in a better ability at holistic, systems thinking. With this, we move on to the second dimension of learning, which is normative.

2.1.3.2. The normative dimension of learning

The normative dimension of learning in a policy context concerns shifts in the norms, values and paradigms that guide the decision-making processes of individuals and collectives (Brown & Vergragt, 2008) In constructivist terms, this kind of learning occurs when claims are 'deconstructed' and lose their factual status through re-identification of their social origins (Jasanoff, 1990, 13). Normative learning is often considered as being of a higher value than cognitive learning, as reflected in terms like double-loop or second-order learning (cf. Argyris & Schön, 1980; Hall, 1993) (Argyris and Schön, 1980; Hall, 1993). Yet as pointed out earlier, this hierarchical notion does not hold in all cases and sometimes inhibits nuanced empirical analysis. We would argue that normative changes are not per se of a higher order, but can

occur at different levels, from the revision of a viewpoint on a particular issue to a fundamental paradigm shift. Normative learning can thus vary both in terms of the abstraction of the norm that is being revised, from very limited and concrete to very far-reaching and broad (this is the logic underlying Hall's learning typology), and its scope and degree of social embeddedness (Schön & Rein, 1995, 33f).

Normative learning occurs through processes of reflection and self-evaluation and therefore has strong links to the literatures on frames, reflexivity and deliberation. Theorists vary in their optimism about the possible extent or means for higher-order normative learning or reframing. A Habermasian perspective and a Foucauldian perspective can be distinguished as two ends of the spectrum. In his theory of communicative action, Habermas depicts the ideal speech situation, 'where public decisions are reached by autonomous citizens in a process of unconstrained exchange of opinion' and where a truthful, free and fair contest of arguments resolves differences in perspective, as key to a democratic society (Habermas, 1984, 102). For Foucault on the other hand, discourses as impersonal systems impose a rigid structure on what can be said and thought (Foucault, 1981), thus severely limiting opportunities for normative learning.

The Foucauldian notion is to some extent present in the literature on frames. At the most general level, frames (Schön & Rein, 1994), belief systems (Sabatier, 1987) or paradigms (Hall, 1993) can be understood as interpretive schemas or clusters of ideas that shape broad attitudes and orientations towards a problem (Surel, 2000). Often, they are tacit and thus escape argumentation and reasoned persuasion. Frames can be seen as processes of questioning 'that structure the world by delimiting the field of possible answers' (Turnbull, 2005, 102), providing individuals with a crucial 'mental grappling hook' (Hoppe, 2010, 181). They are socially co-constructed, i.e. constituted by and changing through the interactions of multiple actors (Putnam & Holmer, 1992; Surel, 2000). Schön and Rein (1994, 26) describe two complementary processes of framing and naming, which allow individuals to selectively single out and 'name' a few key aspects and causal relationships from their complex environment and thereby enable a 'normative leap from data to recommendations, from facts to values, from 'is' to 'ought'.' At the same time, the 'relativist trap' that a frame represents makes frame reflection tremendously difficult; if frames determine what we perceive and how we perceive it, how can evaluation and comparison across frames ever be possible (Schön & Rein, 1994)? Yet as Schön and Rein (1994, 57) themselves admit, 'we know that people do sometimes change their minds, even in fundamental ways, and we know of not-so-rare events in actual policy disputes, where positions have been reframed [...] to open up to accommodation controversies that had at first seemed hopelessly intractable.' Based on their own empirical material, they admit for the possibility of 'situated frame reflection' (Schön and Rein, 1994, 174). They conclude that normative learning may take place rather on the basis of concrete experiences in the context of specific problems and practices than via

reflection about abstract higher-order values with a view to uncover implicit and taken-for-granted assumptions (ibid.). Reframing is thus not seen as a process where one paradigm comes to neatly replace another, but one that initially involves doubt about an existing frame that does not match actual experiences. These doubts then lead to a re-ordering of elements and of new associations, and eventually result in the emergence of a new, distinct paradigm or frame (Kuhn, 1962; Laws & Rein, 2003; Surel, 2000).

Hajer, building on a Foucauldian notion of discourse, but also drawing on concepts from communication theory, puts forward a related, yet distinct perspective. In his view, individuals do not adhere to firm belief systems or frames, but have 'vague, contradictory, and unstable 'value positions' (Hajer, 1995, 71). These can be influenced and fundamentally altered through the emergence of new story lines if these succeed in changing people's views on their own role and on the potential for changes to be made. Story lines are understood as linguistic, multi-interpretable devices which serve to reduce discursive complexity and which bind together discourse coalitions competing for dominance in the policy arena. Discourse coalitions are thus not held together by interests or strong beliefs, but merely by their adherence (for various reasons) to a common storyline (Hajer, 1995). The last point is where Hajer differs from Sabatier's advocacy coalition framework (Sabatier, 1987, 1988, 1998). The advocacy coalition framework is at its core a theory of policy change, but accords 'policy-oriented learning' an important role in this process – and the latter is what is of interest to us here. An advocacy coalition in Sabatier's framework shares a belief in a joint 'policy core'. Learning takes place both within and between coalitions. Importantly, where learning occurs, it is mostly instrumental in nature: 'policy-oriented learning […] is an ongoing process of search and adaptation motivated by the desire to realize core policy beliefs' (Sabatier, 1988, 151). The likelihood of normative learning is inversely related to the degree of commitment to a belief. Learning therefore mostly concerns the secondary aspect of one's belief system (ibid.). Sabatier is very sceptical about the possibility to change the 'deep core' or even the 'policy core' of advocacy coalitions. He does accept the possibility that such a change might happen, but thinks it extremely rare, likening it to a 'religious conversion' (Sabatier, 1998, 104). Peter Hall is somewhat more optimistic in this regard. In his view, learning can extend beyond learning about implementation and how to better reach one's goals to the level of policy goals. In his analysis of economic policy-making in Britain, he identifies a type of third-order learning, which is 'marked by the radical changes in the overarching terms of policy discourse associated with a 'paradigm shift' (Hall, 1993, 279) and encompasses a shift in fundamental beliefs and values underlying public policy (cf. Weiss, 1977). Third-order learning occurs only under special circumstances, and when it happens, usually affects not only governmental circles, but society at large (Hall, 1993).

Other authors in the deliberative and participatory policy analysis tradition attribute a much more central role to normative learning and deliberation in the policy process, a perspective that is embraced in this thesis. Dryzek made an important contribution in this regard and we elaborate some more on his work as it is representative for this line of thinking. As done earlier by Renn et al. (1995), Dryzek translates Habermas' abstract communicative ideal to a more concrete, applied level. The notion of authentic deliberation – which Dryzek defines as requiring communication that induces reflection in a non-coercive fashion (Dryzek, 2000) – is central to his model of deliberative democracy: 'one defining feature of deliberative democracy is that individuals participating in democratic processes are amenable to changing their minds and preferences as a result of the reflection induced by deliberation' (Dryzek, 2000). Dryzek acknowledges the importance of discourses as frames of reference that allow for organising disjointed pieces of information into coherent story lines and allow those who subscribe to a particular discourse to communicate in a meaningful way. Yet, contrary to Foucault, he is convinced that engagement is possible also across the boundaries of different discourses. He believes that individuals are no prisoners of the discourses that they identify with, even if these have shaped their thinking to a significant extent (Dryzek, 2000). Much of the literature on participatory appraisal follows this line of argument (Garmendia & Stagl, 2010; Maarleveld & Dangbegnon, 2002; Muro & Jeffrey, 2008; van de Kerkhof, 2006).

Thus, to summarise, this thesis considers normative learning, which it understands to comprise shifts in viewpoints, values and paradigms that may vary in their degree of abstraction and social embeddedness, a key feature to improving environmental governance. Which factors may facilitate such learning? Not surprisingly, a new and uncertain environment – prompted by a novel issue, a crisis, or severe policy failure – is considered to aid motivation to analyse new information and revise existing viewpoints (Checkel, 2001). In terms of the types of individuals most prone to normative learning, evidence from social psychology suggests that there is a u-shaped relationship between the complexity of one's belief system on a certain subject and the extent of certainty about one's ideas on it. Whereas experts and total novices are generally more reluctant to alter their perception, those with a moderately complex belief system tend to be less sure of their convictions and thus display the greatest potential for normative learning (Chi et al., 1988; cf. Levy, 1994). Finally, experts on participatory appraisal find that the kind of appraisal method matters for the extent of normative learning it provokes among participants. Van de Kerkhof (2004, 79f) argues that those methods involving a high degree of participation (like a policy Delphi exercise or certain applications of simulation gaming) may be more apt to stimulate normative learning than methods that rely on more consultative modes of involvement (which, according to her, may serve better to stimulate creativity and cognitive learning). Finally, and again not entirely surprisingly, a sufficient level of issue-related (and thus not personal) conflict is said to be beneficial for normative learning in stakeholder exercises (Cuppen, 2009).

After this exploration of the two dimensions of learning – cognitive and normative – that form the core of the policy learning literature, we now turn to a main contribution made mainly by writers on social learning. Their focus on smaller social units of analysis has led them to take a more thorough look at the inter-personal, relational dimension of learning.

2.1.3.3 The relational dimension of learning

As noted earlier, much of the literature on social learning developed in the context of natural resource management and environmental appraisal. The focus on relatively small groups of actors who are – or become – familiar with each other and who need to collaborate explains the emphasis that this strand of writings puts on the relational side of learning. The conscious departure from the cognition-focused approach of writings on policy learning is reflected in the quote on the definition of social learning used in the HarmoniCOP project: 'social learning must be conceived as more than just cognitive learning. Learning together to manage together has also to do with changes in attitudes, beliefs, skills, capacities and actions in and among the counterparts' (Craps, 2003, 8). Similarly, Webler et al. (1995) propose to analyse two components of social learning; next to 'cognitive enhancement', studies should also examine the 'moral development' of participants. Whilst the term may smack of normative learning, criteria for the latter category include being able to take on the perspective of others and developing a sense of solidarity with the group (Webler et al., 1995, 446) – issues firmly in the realm of relational learning.

Writers in this strand of literature take significant inspiration from a variety of learning theories from the realms of (adult) education and social psychology. Notions like experiential learning emphasise the direct exposure to the phenomena being studied (Kolb, 1984), similar to the concepts of situated or embedded learning, which focus on the fact that individuals learn best through interactive situations where there is an immediate 'need to know', and in the process engage in and foment 'communities of practice' (Lave & Wenger, 1991). All of them fit the small group or community context in which many of the social learning studies are taking place.

At the most general level, the relational dimension of learning is about how the relationships of participants, but also their sense of self and their identity change in the course of a group process. One of the most important aspects in this regard relates to participants' understanding of each other's viewpoints, arguments and underlying frames. Especially when faced with complex, unstructured problems, stakeholders are often unaware of their own, but also of the wide variety of alternative perspectives that others may have on this issue (Hisschemöller et al., 2001, 62). According to the deliberation literature, as citizens consider the arguments of other citizens, they increase their understanding, empathy with, and tolerance of the variety of perspectives one can have on the public good (Gutmann & Thompson, 1996; Warren, 1996). As Mathews (1994, 235) puts it, 'deliberation doesn't

necessarily change personal positions, but it does change attitudes about opposing points of view.' An orientation towards authentic dialogue as well as mutual respect – enabling the 'constructive interaction [...] with the persons with whom one disagrees' (Gutmann and Thompson, 1996, 81) – may be helpful in this regard. Pearce and Littlejohn (1997) describe dialogue as an orientation to conflict that is open to changing not just what one believes but also how one talks and even thinks about an issue. It involves at least temporarily putting to rest conflict about contrasting worldviews and instead making an effort to collaboratively reflect on them (Burkhalter et al., 2002). One of the outcomes of such an attitude is what Forester (Forester, 1999, 107) terms 'diplomatic recognition': taking the other seriously and treating his concerns as equally legitimate as one's own.

Improved understanding is often expected to result in a positive change of attitude among the individuals involved (Garmendia & Stagl, 2010, 1715) and supposed to lead to the forging of new, or the transformation of existing relationships (Muro & Jeffrey, 2012), including the emergence of new networks and communities of practice. As people develop trust in each other, they embark on a 'virtuous circle of anticipation and action whose initiation always requires a leap of faith beyond the available evidence' (Schön & Rein, 1994, 179). The development of trust is closely related to the emergence of social capital (Ostrom & Ahn, 2003). Social capital, 'networks together with shared norms, values and understanding that facilitate co-operation within or among groups' (OECD, 2001, 41), can be both a driving factor and an outcome of relational learning: 'where there is evidence of ongoing or continuous social learning, then social capital may be produced and/or increased' (Plummer & FitzGibbon, 2007, 56). Finally, stronger ties within a group and increased trust also boost its potential for joint initiatives and collaboration (Armitage et al., 2008; Blackmore, 2007; Forester, 1999; Garmendia & Stagl, 2010). Learning is thus not only at the root of individual behavioural change (although by no means a necessary or sufficient condition for it), but arguably also an important driver for action at the level of the collective. Better understanding of one another may give rise to new opportunities for cooperation that seemed elusive earlier (Schusler et al., 2003).

Relational learning, however, does not only shape relationships, it can also have a significant impact on one's sense of identity. Increased self-confidence or a sense of empowerment are often reported as learning outcomes of participatory processes (Rist et al., 2006), not necessarily in the sense of stakeholders gaining a more optimistic outlook that their voice is heard in the policy process, but in terms of understanding and forming an opinion about complex issues which one previously thought beyond one's reach (Huitema et al., 2010). Relational learning can further touch upon how individuals perceive themselves in relation to the wider community: 'it is in the process of learning about the viewpoints of the others that stakeholders can [...] see how their particular concerns and issues affect, and are affected by, the larger whole of which they are part' (Tippett et al., 2005, 292). The realisation that

there are shared values may give rise to a stronger sense of belonging to political units, as reflected in Webler et al.'s assessment of the impacts of a participatory process on waste siting in Switzerland on the participating stakeholders: 'many learned something about what it meant to be a citizen' (Webler et al., 1995, 458).[6]

2.1.4 Policy-oriented learning in environmental governance – a typology

In developing our own typology of learning in and for environmental governance, we build on the bodies of literature reviewed above, while aiming to address some of the perceived shortcomings of existing conceptualisations of learning types and their operationalisation. We understand policy-oriented learning at its core to be a 'change in understanding', which may – but does not necessarily have to – result in change that is observable to the outside world. The first two learning types in our conceptualisation, cognitive and normative learning, have their roots in the writings on learning in public policy and organisational theory. The last, relational learning, builds on the notion of social learning as it is frequently employed in the domain of natural resources management. We prefer to steer clear of a 'levelled' understanding of learning, treating the different types of learning side-by-side in a non-hierarchical manner. Analysing learning effects in terms of their nature (cognitive, normative, or relational) rather than their perceived value seems both less biased and more systematic from an empirical viewpoint. Table 2 below contains the definition of each learning type in our three-pronged typology. Finally, on the question of the social unit of analysis, given its empirics, this thesis focuses on individual learning.

Type of learning	Definition
Cognitive learning	The acquisition of new knowledge and facts, or the improved structuring of existing knowledge, systems thinking
Normative learning	A shift in views, values, or paradigms
Relational learning	Improved understanding of mind sets of others, enhanced trust and ability to cooperate

Table 2. Types of learning in environmental governance.

Having thus established the notion of learning that will guide the empirical research in this thesis, we proceed to introducing simulation gaming, the appraisal method whose learning benefits we will examine in the following.

[6] Huitema et al. (2007) report similar outcomes of a series of citizen juries organised in the Netherlands. They conclude however that this positive effect depends on the degree to which the authorities take the citizens' recommendations seriously.

2.2 Simulation gaming as a tool to stimulate policy-oriented learning

2.2.1 Complex environmental problems and integrated assessment

Policy exercises, the appraisal method we focus on in this study, belong to the realm of integrated assessment. Since the 1970s, researchers and practitioners alike had become increasingly aware that the complexity of global environmental change might not be usefully captured by strictly disciplinary approaches and methods. This led to the emergence of integrated assessment (IA) – an umbrella term for a range of techniques that, despite many different interpretations and definitions, share two basic elements: interdisciplinarity and decision-support (Rotmans, 1998, 155). The term 'assessment' invokes the objective of producing policy-relevant information, rather than generating knowledge for its own sake. 'Integrated' refers to the fact 'that it brings together a broader set of areas, methods, styles of study, or degrees of certainty, than would typically characterize a study of the same issue within the bounds of a single research discipline' (Hisschemöller et al., 2001, 58). Approaches that would meet these objectives have existed for much longer. As a distinctive strand of research and a community, however, IA developed only from the 1980s onwards. It aimed at filling a gap in two regards: firstly, it sought to account for previously neglected complex feedback and interaction mechanisms in coupled human-environmental systems. Secondly, the objective was to develop frameworks that would allow for testing policy options and trade-offs between various strategies (Rotmans & Van Asselt, 2003, 240). IA is thus often about boundary work between various scientific domains and between the domains of science and policy (cf. Gieryn, 1983).

Looking at the subjects covered by IA processes, past efforts have addressed a range of different functions, from risk and options assessment to goal and strategy formulation (Kloprogge & Sluijs, 2006; Toth & Hizsnyik, 1998). Important advances have been made in IA modelling over the past decades and IA models have occupied a prominent role for instance in assessments of the International Panel on Climate Change (IPCC). Yet they continue to face technical as well as theoretical limitations. Model-based analyses at a large scale, heavily relying on 'economy of representation' (Parson, 1997, 271), cannot expand the range of options under consideration nor can they assess contingencies or strategic or behavioural uncertainty (ibid.). Therefore, model-supported approaches to IA have come to be supplemented with a more participatory strand focusing on the involvement of relevant stakeholders or citizens into the assessment process, through delphi exercises, focus groups, policy exercises or other forms of participatory appraisal (Toth, 2003).

Its proponents claim that participatory impact assessment (PIA) is rooted in the tradition of post-normal science[7] and responds to the perceived need for extended peer-review in the face of the complexities of global environmental change (Funtowicz & Ravetz, 1993; see Chapter 1). It is based on the notion that the inclusion of more perspectives and multiple sources of knowledge serves to enrich the assessment and ultimately results in more robust decision-making (Salter et al., 2010). PIA projects to date have mainly addressed stakeholders, often expending considerable effort to include a diverse set of viewpoints into the process (ibid). In a few cases, most prominently citizen juries, the approach has been broadened to include or explicitly target lay persons as participants. At the same time, it is not unusual for PIA processes to include or even to exclusively consist of scientists. After all, an assessment of policy options requires a critical evaluation of the available knowledge base, and vice versa (Hisschemöller et al., 2001, 63).

Van Asselt and Rijkens-Klomp (2002) map methods for PIA along two axes: the goal of application and the type of targeted output. Regarding the latter category, some methods specifically aim at uncovering and articulating the diversity of perspectives on a particular issue or to test strategies in a safe environment. In others, PIA is a means of achieving closure by reaching consensus among participants on the issue at hand (ibid). Concerning the goal of a PIA process, in some cases, the primary objective may be the process itself and the resulting empowerment of the individuals participating in it. In others, PIA is first and foremost a means to improve the quality of decision-making. However, there are few instances to date where there is evidence that the outputs of a PIA process have had a direct impact on policy-making (Kloprogge & Sluijs, 2006; Salter et al., 2010). This has to do with the issue and difficulties of boundary work; while the results might in principle be valuable to decision-making, institutional obstacles prevent them from being taken up. However, many projects have also been quite methodologically-oriented and experimental in nature, aiming at method development and at improving PIA practice (Salter et al., 2010). Salter et al. (2010) agree with others that one of the most important results in terms of process is the social learning that PIA exercise can stimulate among participants, policy-makers, and the organising team. They emphasise that evidence to this end is hard to obtain and in most instances where it is reported, largely anecdotal in character.

The rationale for involving stakeholders in PIA projects can be either primarily argumentative or cognitive (Hisschemöller et al., 2001). An argumentative approach to PIA focuses on the clarification of stakeholders' perspectives and underlying assumptions through structured dialogue, which is seen as key to integrating

[7] Post-normal science is a term coined by Funtowicz and Ravetz (1991, 1993) with which they aim to characterise a method of inquiry that is suited for cases where 'facts are uncertain, values in dispute, stakes high and decisions urgent' (Funtowicz & Ravetz, 1993, p. 744).

knowledge and to resolving controversies (Cuppen, 2009; van de Kerkhof, 2004). By contrast, the cognitive approach rests on the perceived tendency of policy-makers to neglect long-term considerations and lay knowledge. In this logic, PIA exercises aim to open up policy-makers' perspectives and to encourage unconstrained and creative thinking, often by creating a degree of distance between policy-makers and the issue they are dealing with, for instance through role play, by moving the problem into the future or lifting it to a higher level of abstraction (Hisschemöller et al., 2001; van de Kerkhof, 2004).

While the cognitive and the argumentative approach to PIA are by no means mutually exclusive, the former represents the main rationale underlying the policy exercise approach, the method at the core of the present study. The next section will briefly review the history of simulation gaming, the family of approaches to which the policy exercise belongs. Outside of the domain of the environment, these have had a long history that we will briefly touch upon before going into more detail on the specifics of policy exercises.

2.2.2 Historical overview

The origins of policy games lie in the military sphere, with some authors tracing the origin of modern war gaming as far back as early 19th century Prussia (Caffrey, 2000). After the Second World War, war gaming evolved from a more narrow scope – the 'art of rehearsing for war' (Mayer, 2009, 827), simulating, testing and exploring military postures and strategies in a socio-technical environment – to a more variable strategic tool to aid military decision-making under conditions of complexity and uncertainty. Around the same time, from the 1950s onwards, gaming-simulation methods spilled over from the military domain to related fields like international relations and crisis management.

A growing realisation of the limitations of formalised methods and computer modelling in the 1960s and 1970s led to an increased focus on 'softer' methods for investigating complex socio-political problems, like scenario analysis and the Delphi method (Mayer, 2009). Simulation games were considered particularly useful in this regard, given their ability to combine quantitative and qualitative data (ibid.). Policy gaming may be most suited to environments about which an intermediate degree of information is available. In other words, knowledge about a subject needs to be sufficient for designing a simulation that achieves the necessary degree of representational fidelity, yet the issue is too complex for less costly, time-intensive methods to provide salient insights into options and solutions (Parson & Ward, 1998).

The 1960s also saw first applications of simulation-gaming outside the military-sphere, for business and strategy development. The RAND Corporation played a pivotal role in developing and piloting these methods, including simulation gaming,

which, from the 1970s onwards, started to be applied also to social and environmental issues. A second hub of the development of simulation-gaming was the University of Michigan, where Richard Duke set up a series of simulation games primarily on urban planning issues. Yet these, while focused on policy and planning, were used primarily in a training context, and had only very weak links to policy-making (Mayer, 2009). Overall, the development of simulation gaming can be seen as part of a broader trend in the decision sciences in dealing with complexity, and weighing the opportunities and constraints of formal versus less formal approaches to policy analysis (ibid.).

Simulation gaming had thus expanded its realm of application from a narrow military context to a much wider range of problems (Duke, 1974). This extension seemed to have a certain logic (Underwood & Duke, 1987, 294) as policy games were deemed particularly suitable to deal with crisis situations in strategic multi-actor contexts as well as with situations where there was a major need for institutional reform (Parson & Ward, 1998, 125). Over time, policy gaming applications have taken multiple shapes, from rigid, computer-supported formats and simulations to more flexible 'free-form' games. Policy exercises belong to the latter category.

What characterises a policy game? Authors tend to agree that policy games, and consequently also policy exercises, rely on two basic ingredients: first, participants or participant teams that take decisions independently from one another and without full information (Schelling, 1964) and second, a set of rules and an objective that guides their behaviour (Parson, 1996). There also seems to be agreement that games are a higher-order term than simulations. The latter 'involves the representation of a system or organism by another system or model, which is designed to have relevant behavioral similarity with the original system' (Brewer, 1974, 3). In this sense, all policy games are simulations, yet simulations often lack the requirement of 'independent decision centers' (Schelling, 1964, 31) to be considered as games. Klabbers consequently suggests to do away with the hybrid terminology of simulation-games, gaming/simulation, etc. and to 'call a game a game' (Klabbers, 2009, 459).

According to Parson (1997), policy games draw their usefulness from combining two characteristics, representation and deliberation. The basis for any policy game is a simplified model of reality or the 'system of reference', which, however, retains the main actors, relationships and cause-effect chains (ibid.). Policy games, like game theory and modelling exercises, thus aim at reducing complexity through structured abstraction. At the same time, by relying on interaction between real-life stakeholders instead of agents with predefined preferences, policy games incorporate a strong deliberative element. Participants jointly explore the future 'possibility space', building up a shared understanding of key concepts and searching creatively for solutions (Joldersma et al., 1995). Bots and van Daalen (2007) describe

six different functions that games can fulfil in policy development in the context of natural resources management: 'research and analyze' (games as a laboratory), 'design and recommend' (games as a virtual design studio for policy options), 'provide strategic advice' (games as 'practice ring' for a client), 'mediate' (games as facilitators in a consensus-seeking process), 'democratize' (games as a virtual consultation forum), and 'clarify values and arguments' (games as virtual parliaments). As we will see in the following, policy exercises fit most neatly into the 'design and recommend' (games as a virtual design studio for policy options) category, but may also produce benefits for the other functions.

2.2.3 The policy exercise format and its applications to climate change issues

The idea of a policy exercise was first advocated by Brewer (1986), responding to the need for adapting the method of model-based gaming from a military or business context to the challenges of long-term policy development in complex social-ecological systems. Brewer defined a policy exercise as a 'deliberate procedure in which goals and objectives are systematically clarified and strategic alternatives are invented and evaluated in terms of the values at stake. The exercise is a preparatory activity for effective participation in official decision processes; its outcomes are not official decisions' (Brewer, 1986, 468). Toth (1988a, 1988b) subsequently made an attempt to operationalise the concept and codify its design process. In Toth's words, 'at the heart of the process [of a policy exercise] are scenario writing of 'future histories' and scenario analysis via the interactive formulation and testing of alternative policies that respond to challenges in the scenarios. These scenario-based activities take place in an organisational setting reflecting the institutional features of the problem at hand' (Toth, 1988a, 237). Yet the specific approach and design steps that Toth advocates do not seem to have found widespread application by policy exercise practitioners, or at least this has not been documented (Mayer, 2009).

While the term policy exercise has been interpreted differently and the format has varied from application to application, policy exercises do share some common elements. First of all, the structure of policy exercises is generally less rigid than for other types of policy games. Their primary goal is to advance thinking about unstructured, messy problems for which the set of relevant choices, cause-effect relationships and outcomes is controversial or unclear. Thus, while participants in a policy exercise, like in other forms of games, play roles and are tasked to make plans and decisions under a fictitious scenario, the storyline is often richer and more elaborate. A hypothetical future setting aims at removing participants from their day-to-day reality and to encourage them to think creatively, be open to and search for new insights and ideas (Parson, 1997). Moreover, policy exercises rarely involve fixed scoring systems that reward or penalise actions by participants. Instead, a control team is often used to provide feedback on the decisions taken by the players and to update the scenario in between rounds (Duinker et al., 1993; Parson, 1995; Toth &

Hizsnyik, 2008). Even more important than the interim feedback is the joint evaluation of the dynamics and outcomes of the exercise by participants and the organising team following the event (van Asselt & Rijkens-Klomp, 2002). Finally, given its open format, the right choice of participants may be even more crucial for a policy exercise than for other games. This is often interpreted as implying that participants should have significant expertise and experience in order to be able to contribute meaningfully to the event and to ensure a sufficient degree of behavioural similarity to real policy-making processes (Parson, 1997).

In terms of applications of the method, the Institute for Applied Systems Analysis (IIASA) played an important pioneering role in the late 1980s and early 1990s. Other groups, like the Stockholm Environment Institute, have been active in the field as well. Also in the Netherlands, several 'gaming style' (Mayer, 2009, 839) policy exercises were implemented throughout the 1990s and 2000s. Policy exercises have been used on issues related to environmental policy, health, education, and regulatory policy reform (see Duke & Geurts, 2004, for examples). A number of applications have focused on various aspects of climate policy. Table 3 provides an overview.

When	Organised by	Focus of the exercise
1990	Jill Jäger et al., Stockholm Environment Institute (Jäger, Sonntag, & Bernard, 1990)	International climate policy
1992	Ferenc Toth, funded by the United Nations Environment Programme (UNEP) (Toth, 1992)	Adaptation options in South East Asia under different climate impact scenarios
1993	Jan H. Klabbers et al., Dutch National Research Program on Global Air Pollution and Climate Change (NRP) (Klabbers et al., 1995)	Policy options for climate policy
1995	Ted Parson, Institute for Applied Systems Analysis (IIASA) (Parson, 1995)	Global climate negotiations - mitigation
1997	Bernd Kasemir et al., EU research project ULYSSES (Kasemir et al., 2003)	Potential of venture capital for climate policy
2004	Kate Lonsdale et al., EU research project ATLANTIS (Lonsdale et al., 2008)	Responses to extreme sea level rise in the Thames region
2008	Kurt M. Campbell et al., Center for a New American Security (Burke & Parthemore, 2009)	Security implications of future climate change for the United States
2008/2009	Andrew Jones et al., Sustainability Institute, United States (Jones, 2009)	Global climate negotiations - mitigation
Since 2011	Red Cross Climate Centre (Bachofen et al., 2013; Mendler de Suarez et al., 2012)	Various games on climate risk management, including using climate information, disaster preparedness, and health issues

Table 3. Past policy exercises focusing on climate change issues.

2.2.4 Policy games and exercises and their potential for learning

The rise of policy exercises reflects the hope and belief that such approaches would help policy-makers and scientists better understand real-world processes. The policy games literature accordingly features a great number of claims as to how games would support learning (see also Sections 5.1, 6.1, and 7.1). In the realm of cognitive learning, one benefit that is almost universally mentioned is the potential of policy

games to share and disseminate knowledge. The need for participants to take collective decisions and accomplish tasks under time pressure is said to foster a more intense exchange of information than would be possible in other workshop formats or forms of group work (Parson, 1997). Jones (1964) emphasises the speed with which information initially held by one person enters into common knowledge in a gaming situation. Schelling (1964, 23) explains this as follows:

> *'games are intensely stimulating, people are very active; ideas and conjectures get tossed around and analysed by a highly motivated group of people; a great deal of expertise is collected in a single room [...] people discover facts, ideas, possibilities, capabilities and arguments [...]'*

Knowledge integration is another aspect that is frequently cited. It can take many forms: between different forms of knowledge, between quantitative and qualitative information, between models and strategic or behavioural principles (Parson, 1996b), and between different disciplines and different perspectives on an issue that diverse participants bring to the table (Brewer, 1986; Kriz, 2003; Underwood & Duke, 1987). The process of knowledge integration is said to have one crucial co-benefit, which Duke (1974, 11) describes as 'gestalt awareness', and which, according to Duke and Geurts (2004) constitutes the foremost objective of any policy game. The simulation character of a game supposedly aids participants in grasping the scope and depth of the issue at hand, as well as the different ways of reflecting on it (Schelling, 1964). Games thus are expected to provide a 'potentially powerful variety amplifier [...] to develop systems awareness of the relevant issues surrounding a policy' (Ryan, 2000, 361). Duke considers gaming as the most advanced mode of communication (the title of his 1974 book refers to gaming as the 'future's language'). According to him, gaming enables a 'multilogue' among participants, allowing them to communicate at a sophisticated level and conveying messages that are rich in meaning (Duke, 1974, 55).

Still in the realm of cognitive learning, policy games have been lauded for stimulating creativity among participants and for acting as effective incubators of new ideas and hypotheses. Parson points to the 'pressure, relevance, and distance' inherent in gaming formats as key to bringing this about (Parson, 1996b, 241). The actions and decisions taken by the participants add an element of surprise and unpredictability to the gaming process that may lead to new insights into risks and contingencies, policy options or strategies (ibid.). But games may also help in recognising the 'hidden face in the picture' – insights into process and issues at the periphery of the topic that is being simulated (Schelling, 1964, 30). The results may deliver interesting input and hypotheses for further research as well as practical lessons for policy-making (Toth, 1988a). Yet policy games may not just generate new ideas, they may also allow for interacting testing and experimenting with various options, policy processes, institutional arrangements and actor constellations (Underwood & Duke, 1987; Van der Meer & Geurts, 1995). The notion of the 'safe environment' that a game provides

is of importance here and represents a necessary precondition for experimentation (Mayer, 2009, 825). The possibility of building variations into the scenarios, rules, roles and participants of a game and of comparing the implications of these differences between runs lends more robustness to the futures aspect of analysis through gaming. As time is usually condensed in a gaming situation, the unfolding of a sequence of actions, events and their outcomes can be simulated and later analysed and discussed during the debriefing session (Vissers et al., 1995).

Comparatively less has been written about the normative and relational learning potential of policy games. Although one of the functions of policy games listed by Bots and van Dalen (2007) is 'mediation', overcoming disagreements and consensus-building or shifting viewpoints and values is not something that is much referred to in the theoretical literature on policy games, or listed as an explicit objective in many accounts of past applications of the method. Yet, where a policy game involves the development of new hypotheses or is used to test policy options or processes as described above, this may well involve normative learning for some or even all the participants, aside from the obvious cognitive learning benefits. Moreover, there is the argument that playing out antagonistic roles in a policy game may help participants to clarify underlying arguments and mind sets that otherwise remain implicit in policy discussions (Parson, 1996b; Underwood & Duke, 1987). While this may not yet constitute normative learning as such, it certainly represents an important precondition for it.

In terms of relational learning, it is argued that by adopting a role that is different from their own in reality, participants may gain a better their understanding of their opponents' mind sets, realise what determines their thinking and what constraints they are operating under (Van der Meer & Geurts, 1995). [8] Moreover, the intensely social experience in joint problem-solving that policy games provide supposedly creates 'good fellowship', and 'lead[s] to remarkably good relations afterwards' (Schelling, 1964, 25), even where individuals had fiercely opposed each other over the course of the game. The shared experiences form a common frame of reference that is said to enable participants to communicate more meaningfully after the game (Schelling, 1964, 28).

Overall, strong claims regarding their learning benefits characterise many of the writings on policy games. Yet remarkably few works support these claims by systematic evaluations of actual game runs. We will address this gap in more detail in Chapter 3. For now, we can summarise that the theoretical literature expects policy

[8] This view, however, is qualified by Nemeth et al. (2001), who, based on social psychological research in group processes, find that devil's advocate mechanisms like policy games are less effective in eliciting viewpoints and fostering a frank exchange of arguments than situations where an authentic conflict exists.

games to deliver particularly in the domain of cognitive, but also, especially if participants adopt roles that differ from their own in reality – in the domains of normative and relational learning. Chapters 5, 6, and 7 will examine in more detail to what extent the learning effects from the policy exercise runs evaluated in this thesis conform to these expectations.

2.3 Summary

This chapter argues that the notion of learning in and for environmental governance builds on several bodies of literature, of which the policy sciences, organisational theory, and natural resources and adaptive management are considered most relevant to the present work. Analyses of learning span a range of social units of analysis, from the level of the individual to policy-making systems at large. There are different views in the literature as to whether learning needs to result in an observable behavioural or policy change, and whether only certain types of change qualify as learning. Contrary to many typologies which embrace a hierarchical understanding of learning, this thesis argues that differentiating between cognitive, normative and relational learning may be a more appropriate way of conceptualising learning in an environmental governance context. Learning is conceptualised as a change in understanding which in the first instance occurs at the level of the individual. Cognitive learning is defined here as the acquisition of new knowledge and facts, or the better structuring of existing knowledge, enabling more holistic or systems thinking. Normative learning refers to shifts in views, values or paradigms, which may vary in their degree of abstraction and social embeddedness. Relational learning finally involves a better understanding of others' mind sets, the development of trust and/or a better ability to cooperate.

The second part of the chapter placed the policy exercise, the appraisal method that is at the core of the empirical research in this thesis, into its wider context. Policy games and policy exercises belong to the tradition of integrated assessment. Originally developed in a military context, over time the method has been applied to an ever wider range of problems in strategic multi-actor contexts. The policy exercise, a sub form of the policy game, has often been the method of choice of issues of global environmental change as it responded to the need of adjusting model-based games to the challenges of long-term policy development in socio-ecological systems. Typically, participants in a policy game assume roles and take decisions in a structured setting that represents a simplified model of the system of reference that the game is supposed to represent. There are many claims in the literature as to the great learning potential policy games offer to participants, from improved knowledge integration to stimulating creativity for developing novel policy options, but also normative and relational benefits are mentioned.

3. Measuring participants' learning in policy exercises

This chapter sets out to achieve four objectives. First, it critically assesses the state of the art regarding evaluations of participants' learning effects from participatory appraisal exercises. Second, it presents the efforts undertaken in this study to improve this state of the art, outlining several innovations in the evaluation framework and measurement tools. Third, it introduces the two policy exercises which later serve as empirical cases for our attempts to assess the learning effects of policy games. Finally, it discusses the remaining limitations in terms of research design and measurements.

3.1 The scarcity of systematic attempts at measuring learning in appraisals

Given the cost, time and effort involved in designing, setting up and running participatory appraisal exercises, one would expect a significant degree of evaluation activity to accompany them. Yet while claims are often made about the significant learning potential of such processes for participants (see Chapter 2), the lack of robust assessments in this area is striking. In a systematic review of analyses of social learning in a natural resource management context, Rodela et al. (2012) find hardly any studies in their sample that would qualify as evaluations in terms of research design, whereas quite a number of papers use 'hindsight' in reflecting on the extent of social learning through an appraisal process in an ex-post manner. Blackstock et al. (2007, 729) equally comment that, despite a rise in participatory projects on sustainability and the burgeoning literature on evaluation,

> *'there appears to be no corresponding systematic evaluation of their impact or much sustained reflection on the lessons learnt … Until such literature exists, the assumptions regarding the substantive, instrumental and normative benefits of PR [participatory research, C.H.], outlined above, remain unexplored.'*

Over the past few years, however, some attempts have been made to start addressing this gap. Among the first were van de Kerkhof (2004) and Cuppen (2009). Cuppen draws on pre- and post-measurements using Q methodology to map a change of perspectives over the course of a stakeholder dialogue. Yet both have operated broad conceptualisations of learning and, in Cuppen's case, have not explored learning dynamics in greater, qualitative detail. Schively (2007) and Deyle and Schively Slotterback (2009) have also done work in this area, seeking to measure group learning in participatory planning processes, using pre- and post-

questionnaires. Garmendia and Stagl (2010) adopted a similar approach to evaluate social learning outcomes of sustainability appraisals. Muro and Jeffrey (2012), finally, used a postal survey to assess social learning from two participatory river basin management initiatives, measuring cognitive and relational changes as well as the extent of consensus-building over the course of the process. Especially the endeavours by Garmendia and Stagl and Muro and Jeffrey resulted in relatively sophisticated operationalisations of social learning and have raised the bar in this area. Yet their reliance on only one evaluation method – written surveys – and the lack of pre-measurements in Muro's and Jeffrey's case – leaves them with a less-than-complete picture of learning processes and outcomes as well as their underlying dynamics.

The lack of robust evaluations of learning outcomes from appraisal processes in general is much the same if not more acute for policy games. The lion share of work done on the effects of policy games has taken place in educational contexts, i.e. on games as a teaching tool (Duke & Geurts, 2004; Ulrich, 1997), which present a specific set of issues and challenges. Most accounts of policy games in other settings, with the exception of a few more systematic studies (Duffhues et al., 2014; Mayer et al., 2014; Zhou, 2014; Bekebrede, 2010; Roelofs, 2000; de Caluwé, 1997), present largely anecdotal evidence of satisfaction, drawing on a few quotes of participants or on short ex post evaluation surveys (c.f. Duke & Geurts, 2004), often leaving aside the question of the broader impact of the game. This is also related to the context in which gaming exercises usually take place; as Duke and Geurts (2004, 211) remark, in many cases

> *'research efforts have to be interwoven with the uses of games in applied settings. As a consequence, most empirical research on policy gaming is a compromise and it is often quite limited from a research-design perspective.'*

It is interesting to note that while all the studies referred to above find some evidence for learning in the appraisals they investigated, their results are clearly more limited than those reported in similar analyses drawing only on interviews or participant observation. As Armitage et al. (2008, 87) comment, '[t]hat individual and group learning is slow, inconsistent and/or unexpected […] should come as no surprise, despite the value placed on learning as a goal and a process.' But it may not only be that learning in such exercises or processes is happening to a lesser extent than designers and organisers assume; detecting learning where it occurs is certainly part of the challenge. Crookall and Thorngate (2009, 19), looking at policy games from a learning perspective, frame it in the following way: 'it is easier to 'see' participants' progress than to measure it; we intuitively sense that simulation/ games help people to learn, even if it is difficult to 'prove'.' Such statements, however, highly unsatisfactory from a research perspective, only underscore the need for more

efforts to systematically assess of the effects of these methods, so as to be able to further confirm or qualify these intuitions.

In conclusion, despite some progress there is still a significant gap when it comes to systematically measuring learning effects from appraisal processes in environmental governance. This thesis aims to further conceptual and methodological progress in this area by devising and testing an evaluation framework that captures different dimensions of learning through various instruments. This approach could become a standard to be applied more widely, providing both commissioning agents and designers of appraisals with valuable feedback as to the benefits and limitations of their chosen approaches, also with a view to facilitate cross-case comparison. The next section presents some broader considerations on how to best measure learning, before outlining the evaluation framework developed in this thesis in more detail.

3.2 Tools for measuring learning

3.2.1 Some overarching considerations

What tools to choose to measure learning? Blackstock et al. (2007) recommend a combination of recorded (such as documents and field observations) and reported data (such as interviews and surveys) for this purpose. They further plead for a strong qualitative component in such investigations, in order to allow for the

> *'study of a case in depth and detail, capturing the richness of people's perceptions and experiences in their own terms and developing an analytical understanding through the aggregation of these individual accounts'* (Blackstock et al., 2007, 732).

The evaluation framework devised in this thesis conforms to these recommendations.[9] The measurements, which combine surveys and concept maps with participant observations and interviews, mix recorded and reported as well as qualitative and quantitative data. As stipulated by Patton (1990), this thesis thus makes use of both method and data triangulation. Moreover, it carefully differentiates between different types of learning. In doing so, we break new ground in learning from policy games and other types of appraisals, which mostly relied on just one or two evaluation instruments and which were usually less explicit about exactly what types of learning they were measuring.

[9] See also Mayer et al. (2014) for their comprehensive framework for the evaluation of simulation games.

Another crucial question in the context of evaluating learning from appraisals is the question at which point(s) in time to apply the measurements. Ideally, data should be 'gathered at multiple points throughout the process' (Blackstock et al., 2007, 732) and, most importantly, capture the situation before and after the treatment – in our case, the policy exercise – to strengthen the internal validity of the research. We did this by means of pre- and post-questionnaires as well as pre- and post- concept maps (see Figure 1 for a timeline of the whole evaluation effort). Ideally, there should also be a further evaluative moment at a later stage to account for the fact that some learning might occur only upon reflection, sometime after the exercise. When using recorded measurements, the specific learning effect of the exercise may however be difficult to distinguish from the impact of other experiences at this stage. Therefore, we relied on self-reported data through interviews with participants for this purpose.

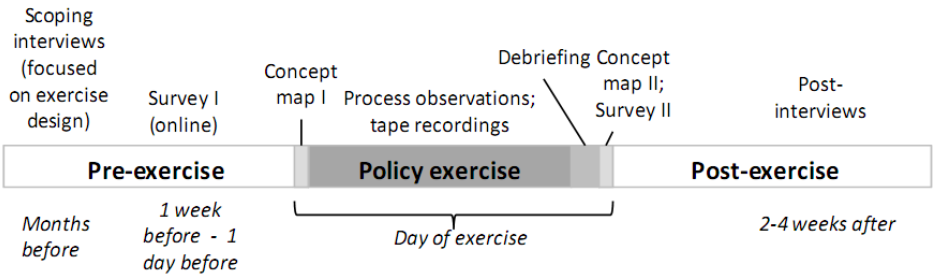

Figure 1. Timeline of the evaluation framework.

A final issue in measuring the learning effect of appraisals exercises, already invoked in Chapter 2, relates to the question of the appropriate level of analysis. While learning occurs through intense interactions in a social setting, measurements are in principle only possible at the level of the individual. This tension is unlikely be resolved completely, but a combination of group-based and individual assessments may go some way in addressing it. In our research design, the debriefing sessions following each exercise allowed for a joint process of sense-making among participants while interviews, surveys, and concept maps captured participants' individual perceptions and attitudes.[10] Protocols from the debriefings have also informed the analysis of learning effects in the subsequent chapters.

[10] As discussed in Chapter 2, the present thesis studies learning at the individual level, and the group means used in the statistical analysis of survey and concept maps are simply aggregations of measurements at the level of the individual. This does not rule out the possibility of additional learning as an emergent property from the collective learning experience participants underwent in the policy exercise, which in turn may have added to learning at the individual level (and hence impacted group means). It is however impossible to filter this out with the research design and methods employed.

Building on these general considerations, the following section describes the different measurement instruments used in this research in greater detail. Table 4 provides an indication of the focus of each instrument with regard to analysing the three types of learning introduced in Chapter 2.

	Cognitive learning	Normative learning	Relational learning
Concept maps			
Interviews			
Surveys			
Participant observations			
Debriefing			

Table 4. Measurement instruments and types of learning. Dark grey shades indicate that a particular tool provided 'important evidence' on a type of learning, whereas light grey shades indicate that a tool served as a 'supplementary data source' that helped confirm or nuance findings obtained from other sources.

3.2.2 The evaluation framework in detail

Concept maps

The most innovative element in our set of instruments to measure learning is the use of concept maps. A concept map is a structural representation 'consisting of nodes representing concepts and […] lines denoting the relation between a pair of nodes' (Ruiz-Primo & Shavelson, 1996, 1). Concept maps, also known as mind maps, are related to other methods such as causal or cognitive maps. They have been used for a variety of purposes, from creative brainstorming to project management, and as planning tools. Yet they also have a history as devices for assessment, especially in higher education (Novak, 1998; Ruiz-Primo & Shavelson, 1996). Various methodologies have been developed for tracing learning through concept maps, among others by Morine-Dershimer (1993).

How were the maps used in the research for this study? Immediately prior to and after each policy exercise, participants were asked to draw concept maps of what they considered to be key aspects of the topic they had been deliberating about, 'burden-sharing in European climate policy' and 'setting up a global REDD fund'. Figures 2 and 3 show the pre- and post-concept map of one expert participant in the exercise on EU burden-sharing. Comparing changes between the pre- and post-concepts maps based on a methodology developed by Morine-Dershimer (1993) allowed to detect how the concepts mentioned by participants gained or lost in importance, or were elaborated in more or less detail from the pre- to post-maps. The method relies on a coding of the items on the maps and is based on two key

principles, (1) centrality – the proximity of certain concepts to the core of the map, which can be taken as an indicator how salient they are in the perception of the author and (2) specificity – the detail with which certain concepts are elaborated on the map, expressed through the proportional frequency with which these occur. A comparison of shifts in the centrality and specificity of concepts from the pre- to the post-measurements therefore allows tracing changes in the structuring of knowledge as well as newly perceived relevance of certain aspects. The assessment methodology is explained in more detail in Chapter 5, which focuses on the cognitive learning effects of the policy exercises.

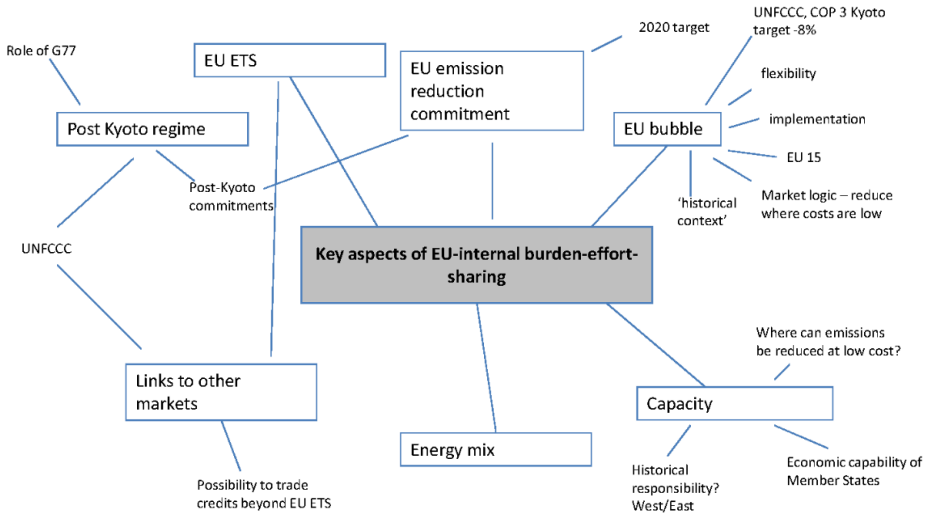

Figure 2. Concept map by one expert participant before the policy exercise on burden-sharing.

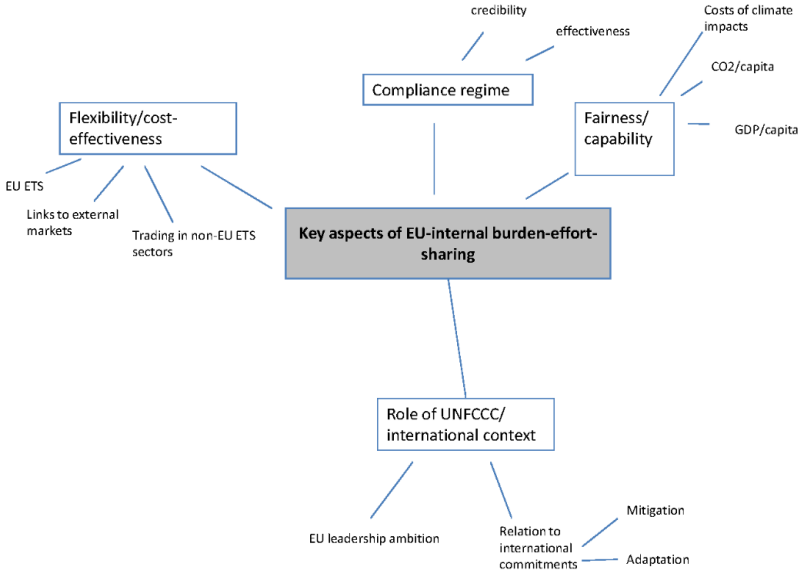

Figure 3. Concept map by the same participant after the policy exercise.

Surveys

Survey data, consisting of a pre- and a post-exercise questionnaire, fulfilled several functions in the research design. Pre-questionnaires, which participants were asked to complete online about one week prior to the exercise, first inquired about participants' professional backgrounds, their motivation for participating in the policy exercise, their knowledge on various aspects of the topic that was being simulated, and if and how they had prepared for the policy exercise.

Beyond this, pre- and post-questionnaires both contained the same set of normative statements concerning, depending on the exercise, either EU-internal burden-sharing or REDD, to which participants were asked to express their agreement or disagreement on a bipolar likert scale. The objective of this part of the survey was to measure normative changes from the pre- to the post-measurement, a technique which has been used in similar settings for instance by Schively (2007), Groves et al. (2008), Deyle and Schively Slotterback (2009), Huitema et al. (2010), and Garmendia and Stagl (2010). Like Schively (2007) and Garmendia and Stagl (2010), we opted for a panel research design, where pre-test and post-test responses of individual participants were matched. While this sacrificed their anonymity, and may have reinforced testing biases like the Hawthorne effect,[11] the possibility to trace individual changes from pre- to post-measurement was considered important enough to warrant this. The normative statements were presented in matrix formatting, but the directionality of their content was deliberately mixed, in order to avoid a 'response set', where individuals tick boxes without properly considering their meaning (Ruane, 2005).

Post-questionnaires further contained a number of self-assessment questions with regard to cognitive, normative, and relational learning through the exercise. Except for an open question, which prompted participants to formulate the three most important insights they were taking away from the exercise, most questions were presented in closed-ended format, to reduce the number of non-responses (Ruane, 2005) and to facilitate subsequent analysis (Czaja & Blair, 2005). Neutral response categories were generally not included in the sets of possible answers, in order to prevent those participants who might not immediately have a strong preference for one of the options from taking the 'easy way out' (Czaja & Blair, 2005, 78). The wording of some survey questions was adjusted slightly for the second policy exercise, based on experience with the first, but remained similar enough to allow for comparison and common analysis.

[11] The Hawthorne effect describes a situation where the participants in an experiment behave differently not as a consequence of the experimental manipulation but because they are aware that their actions are being observed (Draper, 2006).

The data analysis was conducted using SPSS. Apart from a descriptive analysis of the variables, we mainly used analysis of variances to test for shifts in group means from pre-test to post-test as well as for differences among the different cohorts participating in the policy exercise. Moreover, correlation analysis served to investigate the impact of other factors that may have helped or impeded the learning process.

Interviews

Interviews formed an important part of the evaluation framework, and were intended to provide the qualitative richness that Blackstock et al. (2007) called for. By conducting ex post interviews with the participants, we sought to complement the insights on participants' self-reported learning gained from the survey data. Our interviews focused in particular on the domain of relational learning (which, given its ´soft´ nature, was more difficult to explore through other tools), as well as the futures' element of the policy exercise and the perceived novelty and usefulness of its outcomes. A sample interview guide for one policy exercise run is included in Annex 1.A. Since we did not have the capacity to interview all participants after the runs, we took care to interview a set of respondents which was representative in terms of the professional (or, for the students, academic) backgrounds present at the exercise. They took place two to four weeks after the policy exercise, wherever possible in person but otherwise by phone. We considered this an adequate time span to both allow for some reflection after the policy exercise while preventing that other experiences would drown out specific memories of it.

Process observations and tape recordings

As with our interviews, our process observations also concentrated primarily on the relational dimension of the policy exercise. During the exercises with professionals, two members of the organising team took notes on the interactions between participants based on an observation protocol, focusing mainly on the style of interactions and on activity patterns of individual participants. We were particularly interested to what extent participants identified with their roles, how they dealt with the constraints imposed by the scenario, and with the futures context of the exercise in general. The recordings were made with the prior consent of all participants, and the presence of the small tape recorders was quickly forgotten over the course of the day.

Debriefing

At the end of each run, a debriefing session allowed participants to share their experiences and to engage in a discussion of what happened during the exercise, what the implications were, and what they were taking away from it. Typically,

debriefings of simulation games should address three key aspects (Lederman, 1984, 1992): validity (i.e. the extent of relevant behavioural similarity achieved between the exercise and reality), reliability (in terms of the sequence of events and actions taken by participants, and to what extent they were logical and predictable) and utility – i.e. the benefits of the exercise, in terms of learning and otherwise. We dealt with each of these issues in turn, in an open discussion format moderated by the chair of the exercise (for the questions on content) and the facilitator (for the questions on process). The sessions were tape-recorded and notes were taken by the organising team. A guide for a debriefing session is included in Annex 1.B.

In conclusion, we used a wide range of evaluation instruments to assess and cross-validate our three dimensions of learning. The insights gained from the survey data, the concept maps, and the ex post interviews were the most important in informing our analysis of the learning effects from the policy exercise which will be presented in the subsequent chapters. The 'softer' observations on the dynamics of the policy exercise process and the debriefing were however instrumental to yielding additional insights. In the subsequent analysis, we will clearly indicate which evaluation instruments we rely upon. We now turn to introducing our empirical cases, two policy exercises designed to explore two pivotal aspects of international climate policy.

3.3 Research design

3.3.1 Two policy exercises on international climate policy

The set-up of this study follows a multiple-case research design. We draw on a total of four runs of two different policy exercises. For lack of thorough empirical studies on learning from policy games (see previous section), we could not base our case selection on established hypotheses. The themes of the two policy exercises were nonetheless chosen very deliberately. Both deal with topical issues in international climate policy, and both address unresolved, complex puzzles, yet which that are limited enough in scope to allow for advancing knowledge within one day, the duration of the policy exercise. The first policy exercise focused on EU-internal burden-sharing – the question how future greenhouse gas reduction commitments should be allocated among member states of the European Union (EU). The subject of the second policy exercise was the design of a fund to be established under the global climate regime to compensate developing countries for reducing their emissions from deforestation and forest degradation ('REDD').

Policy exercises are considered well suited to addressing institutional questions in complex multi-actor settings as they provide a deliberative yet structured venue for interactions among their participants (Parson, 1996b). Issues relating international

climate policy certainly fit this description and have been addressed in such settings a number of times already (for examples, see Table 3 in Chapter 2). A key aspect that burden-sharing and REDD finance have in common is their strong equity implications: poorer, less developed countries will only commit to climate mitigation if it is adequately supported by their wealthier counterparts. Moreover, the two topics embody central elements of the global climate regime: the burden-sharing issue exemplifies the core of the targets-and-timetables debate, and REDD finance hones in on the discussion on the so-called flexible mechanisms, where industrialised countries pay (or implement projects in) developing countries for reducing their emissions.

While comparable in the significance that issues of international equity take, the two issues differ in one crucial aspect that may also have an effect on the different types of learning that we can expect from the runs: their position in the policy cycle. The stages model of the policy cycle divides the policy process into separate steps of agenda-setting, policy formulation, decision-making, implementation and evaluation. While discredited as empirically inaccurate in its assumption of a linearity of the policy process (Sabatier, 2007), this model still serves as a useful basic heuristic to describe the state of maturity of a policy (Jann & Wegrich, 2007). Following this logic, burden-sharing approaches to climate change mitigation can be considered more mature insofar as such approaches have been discussed and implemented in European – and international – climate policy since the early 1990s. REDD on the other hand has made it onto the agenda of the global climate negotiations only recently and currently finds itself at the policy formulation stage. This difference in maturity of the topic may have implications for the potential for participants' cognitive and normative learning, as we discuss in Chapters 5 and 6.

We held two runs of each policy exercise, one with a group of policy-makers and experts, and one with Master's students in Environment and Resource Management at the Vrije Universiteit Amsterdam.[12] The differences in background between the expert and the student group allowed us to study possible relationships between different levels of experience and expertise and the various types of learning taking place in the policy exercise. The runs with students thus served to strengthen the external validity of the study by expanding the number of observations and repeating the intervention under slightly varying conditions. At the same time, the design of the exercise on EU burden-sharing and on REDD finance was kept as similar as possible in order to keep variations that might have an impact on the dependent variable to a minimum. Both policy exercises were embedded within larger collaborative research projects under the EU Sixth and Seventh Framework Programme. The burden-sharing

[12] http://www.vu.nl/en/programmes/international-masters/programmes/e-f/environment-and-resource-management/index.asp

exercise was organised in the context of the ADAM Project ('ADaptation And Mitigation Strategies: supporting future European climate policy').[13] The REDD-ALERT project ('Reducing Emissions from Deforestation and Degradation through Alternative Landuses in Rainforests of the Tropics')[14] provided the setting for the exercise on REDD financing. More details on the design and set-up of the two policy exercises are provided in Chapter 4.

The topics of the exercises were high on the international political agenda at the time of the expert runs. This, together with the visibility and reputation of the overarching research projects and the fact that we successfully recruited well-known chairs for both events, enabled us to attract two groups of very knowledgeable, experienced mid-career experts and policy-makers to the workshops. Prior to the exercises, scoping interviews with key stakeholders had helped us sharpen the focus and questions to be addressed in the workshops and contributed to identifying key experts and policy-makers to be invited. 22 experts from ten, mostly European countries participated in the burden-sharing exercise. The REDD exercise included an even more international group of 17 experts, with representation from three different continents and a variety of backgrounds. Participant lists for both events are provided in Annex 1.C.

Experts from think tanks and academia made up the single largest group of participants in both policy exercises, with other participants coming from non-governmental organisations (NGOs), consultancy, policy-making, and the private sector. Based on the information they provided in the ex-ante surveys, participants' self-reported level of expertise on the issue discussed at the workshop (European climate policy/burden-sharing and REDD, respectively) varied from medium to very high. Attracting national and EU level policy-makers to the events proved unfortunately difficult despite significant efforts on our part – a challenge which is not unusual for PIA endeavours (Toth, 2003). Busy agendas and the difficulty of convincing higher echelons of the hierarchy of the usefulness of travelling to a workshop that may not be perceived as key for day-to-day work are the likely reasons. In the end, however, a few individuals with long-standing experience in policy-making did participate in both events. In the burden-sharing case, we managed to attract a rather crucial actor in this policy domain, the expert at the European Commission in charge of the burden-sharing portfolio who at the time was managing the negotiation process of the 'effort sharing decision' on EU-internal post-2012 burden-sharing.

For the students, participation in the policy exercise was a compulsory part of the Environmental Policy module in their master's course Environment and Resource

[13] Contract No. 018476-GOCE
[14] Project No: 226310

Management. Students in this course come from a variety of academic backgrounds (ranging from engineering to philosophy) and from a wide range of nationalities. While the Master's students of the 2009/2010 academic year participated in the policy exercise on burden-sharing, those studying for their Master's in 2010/2011 took part in the exercise on REDD financing. In total, 65 students joined the burden-sharing exercise (divided into three different groups over the course of two days), and 86 participated in the REDD exercise (divided into four different groups over two days).

3.3.2 Examining the resulting learning effects

We examined the learning effects resulting from the policy exercise runs from several angles. Because the participants in the two student groups and in the two expert groups can be considered quite similar, we were able to systematically compare learning effects along two axes. On the one hand, we were able to isolate the effects of high (experts) vs. relatively low expertise (students). On the other hand, we were able to compare the impact of a 'mature' policy topic (EU burden-sharing) vs. that of a novel topic (REDD financing). Figure 4 below visualises these two dimensions, along which we have aggregated and disaggregated our empirical data in the following chapters. We also sought to examine the possible impact that a number of contextual factors may have had on participants' learning, such as their professional and academic backgrounds, their reasons for participating in the policy exercise (applicable only for the expert groups), as well as the roles they assumed during the intervention.

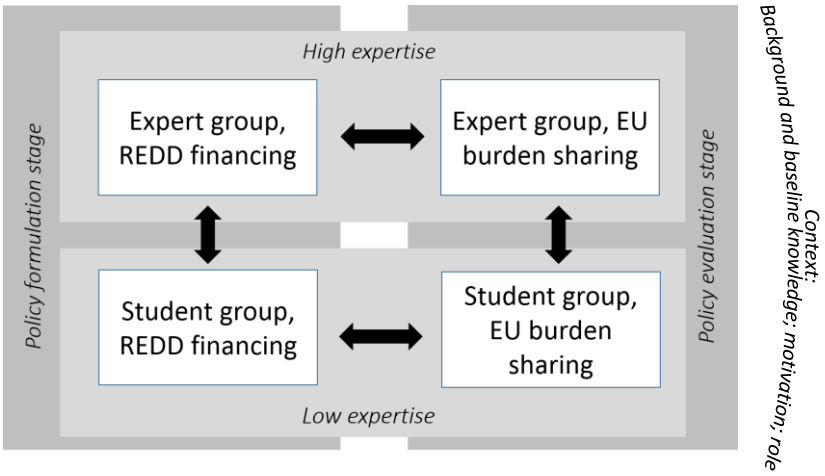

Figure 4. Policy exercise runs and axes along which they are compared.

While the limited existing literature on the learning effects from policy games does not provide any basis for hypothesis-testing, the following table formulates a number of tentative propositions on what we expected participants might learn from the exercises, as well as possible relationships between the analytical matrix sketched out above and the cognitive, normative and relational learning effects from the policy exercises. We will come back to these in the empirical chapters 5, 6 and 7, as well as in the concluding chapter 8 (see Table 22 for a summary of the findings in this regard).

	Expected learning foci within the three learning types	**Expected differences between expert and student groups**	**Expected differences between the more mature policy issue (EU burden-sharing) and the less mature one (REDD financing)**	**Expected relevance of contextual factors**
Cognitive learning	Understanding/awareness of policy options and their feasibility; both for current and future policy; policy dynamics (especially for students)	Higher for students than for experts	Higher for REDD financing (only for expert groups)	Higher for participants with lower baseline knowledge; for experts further removed from policy-making and students with less related academic backgrounds;
Normative learning	Changes in the perception of the desirability and feasibility of policy options and design features; no expectation of normative convergence	Higher for students than for experts	Higher for REDD financing (only for expert groups)	Higher for participants with lower baseline knowledge; higher for roles further removed from reality
Relational learning	Perspectives and interests of the countries represented in the policy exercise	Higher for students than for experts	No differences anticipated	No specific expectations

Table 5. Propositions on the learning effects of the policy exercises.

3.4 Scope and limits of the study

Given the dearth of studies seeking to systematically evaluate learning effects from policy games, the research presented in this thesis is largely exploratory in character. Its goal is to gain insights into a relatively new and little researched phenomenon (Mouton & Marais, 1988; Ruane, 2005). The fact that we have analysed two comparable policy exercises and conducted two runs of each, however, enhances the generalisability of our findings (Stebbins, 2001; R. K. Yin, 2003). Furthermore, the external and internal validity of the research is underpinned by a pre-test–post-test design. Different, alternately independent tools (surveys, interviews, process observations, and concept maps) are used to ascertain that changes in the dependent variable (cognition, attitude, views of participants) are indeed attributable to participation in the policy exercise.

In terms of scope, it is important to underline that the present thesis focuses on participants' learning from the policy exercises. The question to what extent, if at all, the lessons learnt from the expert workshops may have filtered through to policy-making is beyond its scope. Moreover, while the final chapter includes some reflections on the experiences of the organising team in developing and running the policy exercises, this study does not attempt to compile a comprehensive evaluation of the events, beyond the learning effects they generated. The same goes for the lessons that the exercises hold for simulation gaming as a discipline. Finally, while the nearly identical design of the policy exercises on EU burden-sharing and on REDD enhanced the robustness of between-group comparisons of learning effects, it may also limit the generalisability of the findings for policy exercises that are designed differently (although our set-up can be regarded as fairly typical for this type of appraisal).

This thesis also faces a number of limitations with regard to research design and data availability. First, a control group would have strengthened the research design. It would have increased the certainty that the differences between pre- and post-measurements can indeed be attributed to the 'treatment', i.e. participation in the policy exercise. We have sought to compensate for this by taking the pre- and post-measurements as close in time as possible to the policy exercise itself (in order to minimise the effect of other, potentially confounding factors). We also cross-checked them with self-reported information from post-surveys and interviews. Setting up a fully equivalent control group would have been very difficult, given the international provenance of participants and their diverse backgrounds. We initially considered approaching individuals who had been invited but had declined to participate in the policy exercise to partake in a control group. Yet we eventually abandoned the idea as it would have been hard to ensure equivalence between the two groups (given

that the decision to participate was essentially a 'self-selecting' criterion) and also to secure their commitment for the measurements.[15]

One might also have considered comparing the learning effects from the policy exercise with those of another method, for instance a more conventional seminar or a Delphi exercise. While such an approach would have introduced an additional level of variation on the independent variable it would have implied added methodological and logistical challenges (e.g. the difficulty of recruiting two comparable and sufficiently large groups of experts to partake in an intervention on the same topic, as well as timing the two exercises to take place more or less simultaneously). We therefore opted for a more 'within-case' research design with variation in terms of the topic of the exercise and the level of expertise of the participants.

A third, unrelated limitation in the research design is the fact that the organising team was also responsible for the evaluation of the policy exercises, implying the risk of evaluator bias. Finally, the data collected itself has limitations. Measuring learning – in any context – is a complex task, even more so in a setting as ours, where relatively high-powered and knowledgeable professionals come together for a short time span and are required to invest a lot in terms of concentration and input. Self-reported measures in such a context inevitably face biases. Participants may both over-estimate learning in an effort to be courteous to the workshop host and/or underestimate learning effects because of a reluctance to admit to learning in a field where they consider themselves experts. At a more practical level, the surveys and concept maps were a task that came on top of a strenuous workshop, which took these professionals off their busy day-to-day schedule. This had repercussions for the response rate, especially the post-measurements: after a long day, many participants wanted to get home without any further delay. This problem was particularly marked for the policy exercise on burden-sharing: while the response rate for the ex-ante concepts maps was close to 100%, the rate afterwards was considerably lower. In the end, this effectively resulted in comparable ex ante and ex post concept maps from only five participants. For the REDD exercise, the picture was slightly more positive; in total, we received complete sets of pre- and post- concept maps from nine out of

[15] One might also contest that the validity of measurements in the student exercises might have benefitted from exercise participation being voluntary, or against financial reward. However, this would have introduced additional bias in terms of motivation, which could be avoided by the whole class of master students being required to attend the exercise. It was made clear to the students at the outset that the exercise was placed outside the regular curriculum of the course and that while participation in it was mandatory, there would be no grading of the level of participation or of the responses in the survey and interviews, and that these would not be shared with the main instructor of the course.

fifteen participants.[16] For the pre- and post-surveys of both exercises, the responses rates were 65 and 78% respectively.[17] As a consequence, sample sizes used in statistical tests reported in later chapters were quite small in a number of instances, which may have prevented us from picking up medium-sized or small statistical effects in the data.

3.5 Summary

This chapter has argued that, despite the rise of interactive appraisals in environmental governance, systematic evaluations of their learning effects are still largely lacking. The efforts thus far undertaken to address this gap fall short on several counts: the appraisals sometimes do not sufficiently distinguish between different types of learning, but above all they rarely involve a sufficient set of measurement tools and controls to allow for robust findings. The present thesis aims to set a higher conceptual and methodological standard in this respect by devising a comprehensive evaluation framework that captures the cognitive, normative and relational dimension of learning in environmental governance. This framework makes use of a variety of measurement instruments (concept maps, surveys, interviews, participant observations, collective debriefing) employed at multiple points in time. It thereby complies with the standards stipulated in the evaluation literature, which underscore that that assessments of learning should combine multiple methods and data sources and measure learning longitudinally. Two policy exercises on topics in international climate policy provide the empirical case for applying this framework. Addressing the issues of EU burden-sharing and REDD financing, both treat complex, unresolved puzzles that have a strong multi-actor, equity dimension. Both exercises were run twice, once with international experts and once with Master's students. This multiple-case research design allows for comparing learning effects across the different groups, by disaggregating and aggregating the resulting evaluation data along two axes: the maturity of the topic of the exercise (EU burden-sharing being a 'mature' policy topic, whereas REDD financing being more 'novel') and the level of expertise of the participants. Finally, the chapter underlined the limitations of the research in terms of scope and research design. Regarding its scope, the thesis focused firmly on participants' learning, and did not attempt a comprehensive evaluation of the policy exercises in general. In terms of data and research design, the lack of a fully equivalent control group and the limited response rates for the

[16] For the expert runs of both the burden-sharing and the REDD exercise, we excluded two and three individuals respectively from the evaluation as they had been involved into the research process leading up to the policy exercise and thus presented a risk of bias.

[17] The response rate was less of an issue for the student runs as participation in the exercise and the surrounding activities was mandatory for them. The response rates here were over 80%.

post-exercise surveys and concept maps in the expert groups represent the most important constraints.

4. Simulating future climate policy – the ADAM and the REDD-ALERT policy exercise

Whereas the two preceding chapters provided the theoretical and methodological backdrop to this research, this chapter describes the thesis' empirical context and content – the two policy exercises on international climate policy that we organised and ran. The chapter starts out with some general considerations about good practice in game design and the specific approach taken in this regard. It then turns to introduce the context of the topic that was explored in the two exercises, EU burden-sharing and REDD financing. The main part of the chapter then describes the design and set-up of the exercises, reviewing the outcomes and broader findings that emerged from the different runs.

4.1 The game design process

Designing policy games has been described as both an analytical and a creative process (Klabbers, 2006). The analytical aspect relates to the care that needs to be taken to create a model of reality that captures key dynamics of the system of reference in order for a game to yield valid results. Yet the process of scenario-writing and the translation of the system of reference into game elements and paraphernalia also require a substantial amount of creativity on behalf of the game designers.

Good practice stipulates an iterative, multi-step approach to the design of simulation-games, where the specific objectives of the game are gradually specified and the pilot design is tested and fine-tuned before the actual run. (Duke & Geurts, 2004; Duke, 1974; Greenblat & Duke, 1975; Salen & Zimmerman, 2006). The present study followed a five-step procedure to game design advocated by Duke (Duke & Geurts, 2004; Duke, 1974). In total, more than half a year of preparation effort went into the design of each policy exercise, including in-depth background research on their topic, interviews with experts in these areas, and discussions with simulation-gaming experts to test preliminary ideas on game design. After a first preparatory phase, which involved clarifying the context of the larger research projects and the role that our policy exercises would play in them, the next stage was dedicated to problem specification, resulting in a 'narrowly defined concern that the problem is to address' (Duke & Geurts, 2004, 281). This was captured in a concise 'problem statement' that we drew up and discussed within the organising team for both exercises. The subsequent step consisted in establishing clear goals and objectives for the policy game to guide the design process. Advancing knowledge on designing appropriate

policy instruments for EU-internal burden-sharing and REDD financing, respectively, were agreed as the overarching goals for the two exercises at hand. The objective was to explore different options for designing policy on these issues, identify trade-offs and synergies between them, and test their political feasibility. The problem specification stage was followed by a thorough systems analysis of the issue to be simulated. This systems analysis, based on a systematic literature and document review, forms the grid for constructing a model of reality in the game. In our case, we spent several months on desk-based background research into the topic of each policy exercise, studying the policy context and the dynamics surrounding the various policy options under discussion. We also conducted a number of scoping interviews with key experts and stakeholders (10 for the first, and 9 for the second policy exercise). These helped complement the insights gleaned from the desk study, especially regarding key actors and their positions, as well as to update us on the latest policy developments.

The truly creative process commenced in the next step, when the key elements identified in the systems analysis needed to be translated into gaming elements. We had to decide which elements of the systems would be integrated into the scenario and role descriptions and which ones would be left out. Moreover, the design team pondered how and by what means real-life processes could be translated into steps of play and game procedures. A concept report incorporating an overview of the system and a detailed description of the set-up of the exercise provided the basis for elaborating the various gaming elements and for building a game prototype. The prototype was then tested and fine-tuned with the project team before the actual runs with professionals and students. The following two sections introduce the topics of the two exercises, before describing each of them in detail.

4.2 Policy context of the two exercises

Policy exercises aim to shed light on topical, long-term policy or management problems (see Chapter 2). As shown in Figure 4 outlining the research design above, our exercises both focus on the realm of international climate policy, in the first case at the European level (burden-sharing) and in the second at the global level (REDD) .

4.2.1 Burden-sharing as a cornerstone of European climate policy

With the question of burden-sharing, the first policy exercise tackled an issue which is prevalent in global climate policy discussions, but which also forms a key element of the European climate policy architecture: the challenge of distributing greenhouse gas emission reduction efforts among countries, or, in our case, among EU Member States. Under the Kyoto Protocol, the European Community signed up to a joint emission reduction objective, which subsequently needed to be translated into

targets for individual Member States. This process used to been known as 'burden-sharing', but has been relabelled into 'effort sharing' in EU policy discussions addressing this question for the post-2012 period.[18] Burden-sharing poses the crucial dilemma of striking a balance between reducing emissions where they are most cost-effective, and accounting for questions of equity: richer member states might reasonably be asked to shoulder relatively greater costs. Increasing differences in wealth and capacity within an expanding Union on the one hand, and the need for ever more stringent emission reduction commitments on the other suggest that the basic underlying problem of how to allocate burdens is there to stay in European climate policy (Haug & Jordan, 2010).

The scope of the EU burden-sharing mechanism has changed over time. While the 1998 Burden-Sharing Agreement for the period up to 2012 covered emissions from the whole of the European economy, the Commission's set of legislative proposals for the EU climate and energy package for the period 2012-2020 effectively divided the relevant target groups into two parts, depending on whether they were inside or outside of the EU Emission Trading System (ETS). The ETS covers industrial emissions through a 'cap and trade' system, a market-based regime which is principally targeted at cost effectiveness. Under the Commission's new approach, effort sharing only applies to the emissions outside the EU ETS (dubbed 'non-trading sectors'). Together, these non-trading sectors were expected to cover around 60% of EU greenhouse gas emissions.

At the time of the expert run of our policy exercise, the subject was highly topical as negotiations on the climate and energy package were still ongoing, the intended centrepiece of the EU's efforts to cut emissions over the period 2012-2020, which also includes a decision on 'effort sharing'. Our goal, however, was to look even further ahead and explore how this policy domain could evolve after 2020. Crucial questions in the design of burden-sharing arrangements relate to the base year from which emission reductions are accounted for, the criteria on which to base the targets, as well as a variety of issues with regard to implementation, monitoring and enforcement.

4.2.2 REDD financing to compensated developing countries for reducing emissions from deforestation and forest degradation

The main aim of the second policy exercise, which was part of the EU research project REDD-ALERT, was to advance knowledge and stimulate discussion on one key aspect of the current debate on designing a mechanism to reduce emissions from deforestation and forest degradation (REDD) under the global climate regime. The

[18] In the following, we use the terms ´burden sharing´ and ´effort sharing´ interchangeably.

concept of REDD[19] first entered the international climate negotiations at the 11th Conference of the Parties in Montreal in 2005, when a group of countries, led by Papua New Guinea and Costa Rica, submitted a proposal aimed at reducing emissions from tropical deforestation through a market-based approach. The premise underlying the proposal as well as the broader concept of REDD was to provide an incentive to developing countries to reduce their deforestation rates; in the event that they succeeded, they should obtain financial compensation (Humphreys, 2008). Since 2005, REDD has become a crucial issue in the ongoing effort to design a global post-2012 climate policy architecture. Industrialised countries are attracted by its professed potential for cost-effective mitigation (cf. Stern, 2006), although realising it may be less straightforward (Corbera et al., 2010; Gregersen et al., 2010). At the same time, many developing countries, especially smaller ones, consider it a key bargaining chip in their quest to receive climate financing (Haug & Gupta, 2013; Haug & Pattberg, 2008).

After several years of negotiations, there is widespread agreement that the implementation of REDD should follow a 'phased approach'. This would involve moving from an initial stage focusing on preparing for the implementation of REDD ('REDD-readiness') via a performance-based fund to the end vision of a greenhouse-gas-based instrument that 'rewards performance on the basis of quantified forest emissions and removals against agreed reference levels' (Angelsen et al., 2009). Currently, we find ourselves in the first of these three phases, with many initiatives and partnerships ongoing to increase countries' capacities and preparedness for REDD implementation. Various analyses have already sought to forecast the third phase, namely the prospects for, and consequences of, integrating REDD into the global carbon market (Obersteiner et al., 2009; Ogonowski et al., 2007; Stockwell et al., 2009). With the REDD-ALERT policy exercise, the goal was to focus on the second, currently little explored interim phase from about 2015 onwards, when REDD funding may already be results-based, but a fully reliable GHG-based accounting mechanism is still absent. While funding during this phase could also continue to be disbursed through multiple funds, we chose to concentrate on a scenario where the various funding initiatives converge into one global, concerted fund approach.

After this short introduction to their policy context, the following sections will describe the ADAM and REDD-ALERT policy exercises in more detail. The section on the ADAM policy exercise provides a broader introduction to the function of scenarios, roles and policy element cards as well as the structure of our exercises, and the choices we made on these issues. The section on the REDD-Alert exercise,

[19] Over the last few years, the term 'REDD' has increasingly come to be replaced by 'REDD+' or even 'REDD++', denoting either a wider scope for REDD or heightened attention to the importance of social and environmental safeguards in REDD implementation. However, for reasons of simplicity, we refer to 'REDD' throughout this book.

which largely replicated the design of the exercise on burden-sharing, will highlight only those points that were specific to that event or that differed from the first exercise.

4.3 The ADAM policy exercise on burden-sharing in EU climate policy

4.3.1 Design of the ADAM policy exercise

As argued earlier in Chapters 2 and 3, the nature of the topics to be explored – design and feasibility of future policy in complex multi-actor settings beset with multiple uncertainties – made the policy exercise format a natural choice for the two interventions. Policy exercises are flexibly structured interfaces between scientists, experts and policy-makers, aimed at eliciting and synthesising relevant knowledge from a variety of disciplines (Toth & Hizsnyik, 2008). Typically, they draw heavily on scenario development and analysis, making use of participants' experience and expertise in exploring future policy choices (Toth & Hizsnyik, 2008).

The aim of the ADAM policy exercise was to appraise options for future EU burden-sharing arrangements in the medium term. Over the course of one full working day and under the guidance of a skilled facilitator, participants were tasked to negotiate the features of a burden-sharing agreement that would allocate emission reduction commitments among EU member states for the period beyond 2020. The exercise was presented in the form of a simplified EU policy-making cycle, with scenarios, role descriptions and sets of 'policy element cards' as key inputs. The policy exercise was preceded by a dinner on the eve of the workshop, during which the basic approach and objectives were outlined. Moreover, one of the architects of the previous EU burden-sharing agreement - a former chief climate negotiator of the Netherlands - gave a talk on his experience at the time, and participants had the chance to get to know each other.

Scenarios

The purpose of a scenario in a policy exercise is to provide a specific background and context for teams' decision-making and to focus attention on the key issues, problems and decisions to be highlighted for examination (Parson, 1995). A scenario should therefore provide enough information to frame the discussion, but should avoid overwhelming participants with excessive detail and over-determining the outcomes. During a policy exercise, participants may walk forward or backward in time through one or more scenarios (Toth, 1988a). In deciding on the number of cycles or scenarios, there is a trade-off between exploring and comparing different alternative 'futures' (or moving through the system in several steps) and the time available for one cycle and decision-making within it.

Given the limited duration of the policy exercise (one working day), it was decided to probe more deeply into one single scenario set in 2018 instead of walking through several scenarios over time. In order to explore how different political and economic circumstances could affect the outcomes of the simulation, participants were divided into two subgroups that went through the same steps of play yet were confronted with two scenarios that differed in some crucial aspects. Thus, the scenario in the first group reported a high degree of international cooperation on climate change under a successful successor agreement to the Kyoto Protocol. In the other scenario, emissions abatement was occurring through a variety of uncoordinated bilateral and multilateral initiatives and efforts at various levels (see Box 1 for a summary, and Annex 1.D for the full scenarios).

The **'Kyoto to Kingston' scenario** was set in the year 2018 and reported the conclusion of a post-2012 climate agreement, in Kingston, Jamaica in 2010. It assumed that the international climate policy landscape had changed dramatically, with the United States as a new green climate policy leader, and significant mitigation action taking place in the BRIC countries. In the EU, by contrast, the 2009 climate and energy package had not delivered on its promise, with effort-sharing in particular proving largely ineffective and controversial among member states. Devising an ambitious yet realistic mitigation strategy for the post-2020 era was thus of utmost importance for the EU to regain its international leadership role on climate change. At the European Council meeting that the exercise sought to simulate, a start was to be made with the conclusion of an ambitious and effective post-2020 burden-sharing agreement.

The **'Coat with many colours' scenario** differed from the above in two key aspects. It assumed that a post-2012 protocol had been concluded in Copenhagen, but had failed to attract sufficient ratifications. In its place, a multitude of multi- and bilateral initiatives had sprung up, with the UNFCCC having transformed into a clearinghouse. EU-internally, while the Kyoto to Kingston scenario had reported significant deficiencies in the functioning of the EU ETS, the assumption here was that the system had finally started to live up to its promise.

Box 1. Summaries of the two scenarios of the burden-sharing exercise.

Roles

In deciding about roles and role descriptions, designers of a policy exercise typically balance two aspects: on the one hand, roles serve to create distance for participants from their day-to-day work so as to better enable them to examine broader, long-term strategic issues and to look at the wider picture. On the other, roles should maintain key aspects of the real-life position and institutional constraints that actors are subject to, so as to not render the policy exercise totally irrelevant (Toth, 1988a).

For the EU burden-sharing exercise, participants were divided into five teams within each of the two scenario groups. The teams represented senior policy-makers of four EU member states (Germany, Poland, Spain and Sweden) and the European Commission. In order to allow for deliberation inside the country teams and to keep the negotiations manageable for a one-day workshop, only a limited number of countries were represented in the exercise. At the same time, these actors were carefully chosen so that they would preserve the representation of the key interests and conflict lines on EU burden-sharing within the Union. Before the policy exercise, participants received short role descriptions summarising important developments in their country's climate policy and economic outlook over the last decade (see Box 2 and Annex 1.E). This information allowed them to deduce the issues that had priority for them in the negotiations. Moreover, a professional journalist published regular news bulletins' on progress in negotiations and on positions adopted by Parties.

Germany in 2018: Robust economic growth over the last decade. Climate policy remains a priority. More or less on track in effort-sharing domain, relying on external carbon credits. Currently holds EU presidency, thus keen to secure an ambitious agreement.

Sweden in 2018: Impressive climate policy record over the last decade, one of the rare success stories in effort-sharing. Yet mitigation potentials increasingly exhausted. Population remains supportive of climate policy, yet expects other countries to step up their efforts as well.

Spain in 2018: has suffered a severe economic downturn. Climate policy has taken a backseat among more pressing concerns. Not on track in effort-sharing domain, but large abatement potentials are still available. Impacts of climate change increasingly felt in agriculture and tourism.

Poland in 2018: has seen high economic growth between 2008-2018, but remains a 'problem child' for EU climate policy. Currently pending infringement proceedings for non-compliance in effort-sharing domain, exacerbated by cold winters in 2015/16. Argues that in hindsight effort-sharing target was excessively strict.

EU Commission in 2018: after the success of the 2009 adoption of the EU climate and energy package, increasingly strained relations between member states and the Commission on climate change issues, due to a number of reasons. Keen on an ambitious agreement that projects the EU's international leadership and that allows for transparent management and accounting internally.

Box 2. Summary of role descriptions for the EU burden sharing exercise.

A crucial choice facing game designers is whether participants should play roles that are close to their own in reality or whether one should encourage role switching. One might call this dilemma one of dramatic distance. Roles approximating participants' real-life occupations ensure that they are played realistically and that all relevant

knowledge is brought to the table. Yet they may also prevent a fresh look at the issues at hand, thereby reducing the potential for normative and relational learning, which might be stronger when participants find themselves in unfamiliar roles. At the same time, the benefits of role-switches are somewhat contested in the social learning literature. Nemeth et al. (2001) find that they suppress authentic and constructive conflict by unrealistically assuaging opposing viewpoints. Our policy exercises opted for a middle ground in this regard. In assigning the roles, care was taken to as far as possible match participants' own backgrounds with the nationalities they were representing. However, as the professional occupations of the participants matched the profile of actual policy-makers and negotiators in only few cases, a significant degree of dramatic distance was maintained in all instances.

Policy element cards

How to frame and bound the subject matter to be examined over the course of a policy exercise is another issue requiring careful reflection. In addition to scenario and role descriptions, 'policy element cards' served to provide structure and focus to the discussions. The coloured cards contained different options for the central features of an EU-internal burden-sharing agreement which participants were asked to negotiate over the course of the exercise. The options were developed based on the various legislative proposals on EU effort-sharing that were being discussed in Brussels at the time as well as position papers and ideas advocated by think tanks and NGOs, not only in the EU, but also in the international climate policy context. Participants were free to amend cards or invent new options on blank cards however they saw fit.

In total, there were seven design features of a burden-sharing agreement on which a decision was required as well as the option to suggest additional EU policies for the effort sharing domain (see Annex 1.F for the complete set of options):

1. *Criteria:* Which criteria or indicators should form the basis for allocating emission reduction commitments among member states?
2. *Base year:* From which reference point should emission reductions be measured?
3. *Limits to external credits:* What share of their targets should member states be allowed to meet through external carbon credits?
4. *Quality of external credits*: What should be the quality requirements for such credits?
5. *Compliance and enforcement:* What should be the rules in case of non-compliance and who should enforce them?
6. *Internal flexibility:* How much flexibility should there be inside the effort sharing domain, i.e. in terms of trading of excess reductions between member states, banking and borrowing of emission reductions ?

7. *Interaction with the EU ETS:* What should be the rules for external flexibility, i.e. the interaction of the effort sharing domain with the EU ETS?
8. *Proposals for new policy initiatives [optional element]:* What additional EU-level policies and measure should be initiated for the effort sharing domain?

The focus on analysing the design features of the burden-sharing agreement rather than the quantitative emission reduction targets of member states was chosen deliberately. It was based on the concern that otherwise, haggling over numbers grounded primarily in scenario assumptions would quickly come to dominate the exercise and yield only limited useful insights. Initially, the organising team considered underpinning the exercise with real-time model runs, for instance based on the FAIR model developed by the Netherlands Environmental Assessment Agency. However, experience from a previous policy exercise prompted us to opt for a purely qualitative exercise. In that case, understanding the assumptions of the model had taken time away from deliberations, and concerns about the models' reliability had been used strategically in the negotiations (Parson, 1995),

Steps of play

The policy exercise consisted of three different phases, simulating a simplified EU policy-making cycle (see Box 3). In the first phase, country teams were tasked to develop their national position on burden-sharing, while the European Commission, reflecting its legal right of initiative, was asked to prepare a proposal as the basis for later negotiations. Initially, to help participants get into their roles, each country team was to choose up to five 'success criteria' from a list that in their view reflected best what their team would consider a successful agreement on EU burden-sharing. Then, teams began with their main task for the morning session, which consisted of selecting their preferred option(s) from the set of 'policy element cards' (or inventing their own alternatives) for each design feature of the post-2020 EU burden-sharing agreement. The selected cards were glued on a cardboard template, with a justification added for each of them. Finally, in a last step, country teams were asked to rank which features they intended to prioritise in the negotiations. Throughout this first block, country teams were free to consult with one other, as well as with the European Commission.

Structure of the ADAM policy exercise:

1. **National Decision-making.** Country teams define their national position on burden-sharing.
2. **Commission presents its proposal** for a burden-sharing agreement to country teams.
3. **Plenary negotiations** based on the Commission's proposal.

Box 3. Steps of play

The next phase saw the Commission present its proposal to the country teams. The country teams and the journalist then had an opportunity to ask for clarifications or

pose questions before the group moved into negotiation mode for the afternoon session. This final plenary session was chaired by Germany, identified in the exercise scenario as the EU presidency in charge of leading the negotiations. The task in the plenary was to reach an agreement on all features of the burden-sharing agreement. Country teams first commented on the Commission's proposal and proposed amendments by attaching post-its to the Commission's proposal posted onto a wall. Then the actual negotiations started, alternating between plenary negotiations and informal discussions in smaller groups. Throughout the process, participants were, aided by the policy element cards, encouraged to focus on ideas rather than negotiation text. In the end, reflecting the rule of qualified majority voting in place in the EU, either consensus by all four countries or by three out of four countries and the Commission was required to adopt the agreement. After agreement had been reached, the last step of the exercise had country teams return to the success criteria they had selected at the beginning of the day. They evaluated to what extent the agreement met their objectives, thus already preparing for the discussions during the debriefing.

4.3.2 Runs and outcomes of the ADAM policy exercise

As discussed in Chapter 3, two runs were held of each policy exercise, first with a group of international experts, and then with a class of Master's students in Environment and Resource Management. We do not discuss the game runs with the students at much length here. The reason is that the runs with students primarily served to measure the effects in terms of the different learning types, and to examine whether and how these differed from those in the expert groups. The substantive outcomes of these runs, by contrast, are far less interesting because they could not fully incorporate and do justice to all the options that arose and because outcomes were far more reflective of particular students' motivation and negotiations skills than was true for the expert run.

Run with experts

The expert run of the ADAM policy exercise was held on 28 October 2008, involving 23 international participants. They were tasked to negotiate a burden-sharing agreement covering eight design features (see Tables 6 and 7). The two sub-groups who played the exercise in parallel faced two differences in terms of scenario design (group A's 'Kyoto to Kingston' scenario featured a unified global climate regime combined with an underperforming EU ETS for; group B's 'Coat with many colours' featured a fragmented climate regime combined with a strong EU ETS). As the day went on, game dynamics evolved quite differently in the two groups. In the morning, the groups still followed largely similar patterns. After a first phase where teams worked individually to make sense of the game material and its implications for their country's position on burden-sharing, interactions between teams increased in the second half of the morning when countries started to consult with their counterparts

and with the Commission. In both groups, this was initially prompted by the European Commission, who sought to collect countries' views in preparation of the draft agreement that it was due to present by mid-day.

The draft effort-sharing agreement that the Commission eventually presented in Group A followed a slightly different logic than the proposal put forward in Group B (see Table 6). By departing from a wealth-based distribution of mitigation burdens and by choosing an earlier base year, the proposal in Group A placed more emphasis on equity considerations and on past performance records of member states than the one in Group B. The quality requirements and limits for the use of external carbon credits were more stringent as well, emphasising domestic reductions over flexibility mechanisms. In Group B on the other hand, the Commission team emphasised maximum flexibility in the effort-sharing domain, by allowing for the trading of excess reductions among member states and by proposing rather generous rules on external carbon credits. Overall, the two Commission proposals mostly used the pre-prepared options on the policy element cards that participants had at their disposal. Only in the subsequent phase did groups start to rely more on their own ideas in amending and combining options.

	Commission proposal Group A, 'From Kyoto to Kingston'	**Commission proposal Group B, 'Coat with many colours'**
Criteria for effort-sharing: *Which criteria or indicators should form the basis for allocating emission reduction commitments among member states?*	Based on relative GDP per capita of member states	Aiming at converging per capita emissions in effort-sharing sectors by 2030, GDP-adjusted
Base year: *From which reference point should emission reductions be measured?*	Average yearly emissions between 2004 and 2006	Average yearly emissions between 2012-2016
Internal flexibility: *How much flexibility should there be inside the effort sharing domain, i.e. in terms of trading of excess reductions between member states, banking and borrowing of emission reductions ?*	MS shall comply with interim targets set for 2025 and the end target 2030.	A member state can trade excess reductions in the effort-sharing sector with other member states Member states shall comply with their targets in a linear manner
Quality of external credits: *What should be the quality requirements for such credits?*	Only credits with Gold Standard or equivalent quality standard are admissible	All types of credits recognised by the UNFCCC are eligible.

Limits to external credits: *What share of their targets should member states be allowed to meet through external carbon credits?*	Yearly quota of 10% per member state of external carbon credits; remaining non-used quota may be passed on to other member states	Limit of total effort 30%
Compliance and enforcement: *What should be the rules in case of non-compliance and who should enforce them?*	Authority administering compliance regime: Commission	Authority administering compliance regime: European Environment agency; infringement procedure under EU law applies
Interaction with EU ETS: *What should be the rules for external flexibility, i.e. the interaction of the effort sharing domain with the EU ETS?*	No exchange between EU ETS and effort-sharing domain.	No exchange between EU ETS and effort-sharing domain. Coverage of EU ETS extended to agriculture.
New EU policy initiatives: *What additional EU-level policies and measure should be initiated for the effort sharing domain?*	Vision 2030; 'DGTR' – Distributed Generation for Transport and the Residential Sector	Stimulate modal shift in the transport sector. Technology investment funds.

Table 6. Draft effort-sharing agreement proposed by the Commission in the two groups.

Outcomes of the expert run

Comparing the outcomes from the two groups (see Table 7 below), there are a number of striking similarities. Both agreements accept all internationally recognised types of external credits, with limits of three per cent of a country's national emissions in the base year for Group B, and ten per cent for Group A, respectively. In both cases, the Commission remains in charge of the non-compliance procedure, and penalties are levied on offenders – an outcome which, interestingly, provides more ´teeth´ to the agreement than the original proposals by the two Commission teams. Both groups opt to keep the EU ETS and the effort-sharing domains relatively separate, even if the final compromise in Group A allows for a 'limited' exchange, without specifying what exactly this entails. Slight differences concern the issue of internal flexibility: only Group B permits trading of excess reductions among member states, but both groups endorse interim targets before the binding endpoint of the agreement in 2030 (in the form of the requirement of a linear reduction path in one case, and of binding interim targets in the other). Furthermore, concerns about the costs of adaptation to climate change have found their way into the agreement in Group B: penalties for non-compliance are to be used for adaptation needs, and the Commission is requested to elaborate proposals for the earmarking of auctioning revenues for adaptation, as well as proposals for improved adaptation policies in general. In both groups, the question of criteria for burden-sharing constituted a

major bone of contention in the negotiations. Group A eventually chose relative efficiency of emissions per unit of economic output as the key criterion,[20] arguing that this rendered the choice of a base year irrelevant as the indicators were to be continuously updated. Group B settled on a combination of criteria to satisfy the dual objectives of cost-effectiveness and equity. Their indicator was composed of the marginal abatement costs per member state (accounting for 50%), GDP per capita (30%) and emissions per capita (20%). When proposing this, the German presidency argued that trends in marginal abatement costs were to be reviewed in 2024, and that, assuming these converged across the Community, emissions per capita would become the most important criterion in the longer term.

	Final agreement on post-2020 effort sharing **Group A, 'From Kyoto to Kingston**	**Final agreement on post-2020 effort sharing** **Group B, 'Coat with many colours'**
Criteria for effort-sharing	Relative efficiency of emissions per unit of economic output.	Combined indicator, 50% based on marginal abatement costs, 30% on GDP per capita, and 20% on emissions per capita. Review of trends in marginal abatement costs in 2024.
Base year	*[Considered irrelevant, given the criterion chosen, see above]*	Average yearly emissions between 2014-16.
Internal flexibility	Interim targets set for the years 2023, 2026 and 2028, but only the latter two are binding.	A member state can trade excess reductions in the effort-sharing sector with other member states Member states shall comply with their targets in a linear manner.
Quality of external credits	All types of credits recognised by the UNFCCC are eligible.	All types of credits recognised under the Kingston Protocol are eligible.
Limits to external credits	Yearly quota of 10% per member state of external carbon credits; remaining unused quota can be banked or passed on to other member states.	Yearly quota of 3% of external carbon credits per member state; remaining unused quota can be banked or passed on to other member states.

[20] Relative emission efficiency per unit of economic output appears to refer to what is commonly understood by carbon intensity - the amount of emissions of carbon dioxide per unit of GDP (Allwood et al., 2014)

Compliance and enforcement	Authority administering compliance regime: Commission. Fines to be deducted from member states' allocations from EU structural funds.	Authority administering compliance regime: Commission. Fines equivalent to penalty payments under the EU-ETS, revenues to be used for adaptation measures (Commission to present proposal).
Interaction with EU ETS	'Limited' exchange between EU ETS and effort-sharing domain.	No exchange between EU ETS and effort-sharing domain. Commission to present assessment of the feasibility of including the transport sector into the EU ETS by 2019, as well as possibilities for earmarking auctioning revenues for adaptation measures.
New EU policy initiatives	Commission invited to present proposals for new policy initiatives and commitments in the transport and building sectors.	Commission to assess to what extent existing cohesion instruments and funds contribute to low-carbon society, and on that basis develop proposals for new policy initiatives, as well as improved adaptation measures. In doing so, should pay special attention to the building sector and to the needs of countries with low GDP.

Table 7. Agreements reached in the two expert groups.

These outcomes were considered by and large realistic by participants during the debriefing. They added, however, that including more quantitative information (such as assumptions about the EU's and other countries' mitigation targets, energy trends, prices, and abatement costs) in the scenarios and role descriptions would have further increased the plausibility of the policy exercise. Participants however expressed support for our decision to run the exercise without modelling support, sharing our view that this would have distracted from the core of the discussions. One aspect that was discussed at some length was the relative similarity of the outcomes reached in the two groups, despite the differences in the underlying policy scenarios, a result that neither the organising team nor the participants had expected. The latter attributed this mainly to the way in which the exercise was set up. The fact that only EU member states were represented and that these were never confronted with the consequences of their actions in the international arena may have prompted a rather introspective, EU-internal outlook, in which the differing international

context did not play a major role in decision-making. We had asked the professional journalist that played the role of the press to emphasise the global dimension by publishing international news items and by questioning participants on the international implications of their decisions. Yet this turned out less effective than hoped. The journalist was soon overlooked in the heat of the negotiations, also because her activity carried no direct consequences for the game.

Based on the debriefing and the ex-post survey, participants were generally quite positive about the event. In the ex post interviews, participants stated that the policy exercise had created an atmosphere 'full of fun and respect' in a very active workshop format that made it impossible for participants 'to stay out of the discussions or give very little input' and which meant that everyone was 'forced to share their expertise'. A point of criticism, however, related to the policy element cards. While it was felt that they had been useful for structuring discussions, participants argued that fewer design features and a smaller set of pre-prepared options might have allowed for more in-depth discussions and might have stimulated participants' creativity even more. The final Chapter 8 reflects in greater detail on the strengths and shortcomings in the design of the policy exercise.

The student runs

Given the large number of master students registered for the module that included the policy exercise, we divided the group and ran the exercise twice on two different days in February and March of 2010. The first run involved around 40 students playing in two sub-groups, identical to the set-up of the expert run. On the second day, around 20 students simulated the 'From Kyoto to Kingston' scenario only. In general, the students coped well with the complex material they were confronted with. Teams familiarised themselves relatively quickly with the scenario and role descriptions and started devising their strategy. Yet particularly in the negotiation phase in the afternoon, it was noticeable how much process and outcomes of the run were determined by the individuals taking part in it. Thus, whereas the 'Kyoto to Kingston' group on the first day negotiated most of the time in plenary and sought to resolve differences in viewpoints in an orderly, bureaucratic fashion, discussions in the 'Coat with many colours' group evolved mainly in small caucus settings, with parties attempting to strong-arm their opponents and deliberately spreading misinformation. The outcomes reflect this to some extent; for instance, due mainly to the negotiating skills of the Swedish team, the goal of economically efficient abatement largely prevailed over equity considerations in the agreement reached in the 'Coat with many colours' group. The Spanish and Polish teams realised only during the debriefing that this implied that they would have to carry the brunt of the burden in greenhouse gas abatement. Thus, whereas according to the debriefing, students considered the exercise valuable part of their curriculum, the limited background knowledge of the groups also meant that the negotiation outcomes were

not exactly realistic, nor had they the potential to produce insights relevant to future climate governance.

4.4 The REDD-ALERT policy exercise on REDD financing

4.4.1 Design of the REDD-ALERT exercise

For the exercise on REDD financing, we sought to replicate the set-up of the policy exercise on EU burden-sharing to the greatest possible extent, in order to minimise variations in the learning effect of the two policy exercises stemming from design differences. Moreover, the aims of the two exercises, the decision-making context, and actor dynamics of the two topics that we simulated were quite similar. The REDD workshop however deviated from the first exercise in relying on a single scenario. During the debriefing sessions of the burden-sharing exercise, experts and students alike had commented that they did not feel that going through two scenarios in parallel had yielded any insights into possible impacts of differing international contexts for EU policy developments. This view was largely shared by the organising team; Chapter 8 reflects further on this issue. As a consequence, we abandoned the second scenario/ two group approach for the REDD-ALERT policy exercise.

As with the burden-sharing exercise, the REDD-ALERT policy exercise lasted for one full working day, preceded by a workshop dinner. Its overall objective was to evaluate and test the feasibility of various options for designing a global REDD fund that would compensate developing countries for action on deforestation and forest degradation that would result in emission reductions. Again, policy element cards served to structure the discussions to this end. In total, eight design features of a global REDD fund were to be negotiated (see Annex 1.H for a full overview of all policy options):

1. *Capitalisation and replenishment:* Where should the finance for the fund come from?
2. *Accounting for co-benefits:* Should high(er) social/environmental co-benefits of REDD activities be rewarded, and if so, how?
3. *Eligibility criteria:* How should specific funding needs be prioritised in view of overall funding constraints?
4. *Fund administration:* Who should manage the fund?
5. *Verification:* Who should verify that REDD activities are realised and result in actual emission reductions?
6. *Payment size:* How should the level of payment to REDD countries be determined?
7. *Modalities for upfront financing:* Should REDD countries have access to upfront financing and if so, how should this organised?

8. [Optional] Recommendation on the *transition of REDD into the global carbon market*.

While figures and precise amounts to be committed to the fund would usually feature prominently in this kind of negotiations, we avoided placing much emphasis on them (just as in the ADAM policy exercise) as we doubted that haggling over fictitious scenario assumptions would yield many useful insights. Instead, we focused on those qualitative design elements that would have an important bearing on the fund's functioning.

We retained the structure of the ADAM exercise and divided the day into three phases – national decision-making, presentation of a proposal for a global REDD fund, and negotiations in plenary. Mirroring the ADAM exercise design, the REDD exercise included five country teams and a journalist representing the press. It was framed as a pre-meeting to the 2015 UN climate summit in Santiago de Chile. The scenario (see Box 4 for a summary, and Annex 1.G for the full version) specified that Brazil had called together a small group of crucial actors (the United States, the European Union, Brazil, India, and Cameroon) to develop a joint proposal for a REDD fund that could subsequently guide the negotiations in Santiago de Chile. Our decision to frame the exercise as a pre-meeting instead of the actual negotiations reflected the limited number of parties that could be credibly represented in our simulation. It would have been impossible to capture the diversity of interests present in UN plenary negotiations, especially since we wanted participants to play in teams to increase the depth of deliberations informing countries' positions, and to maintain the similarity to the set-up of the ADAM exercise.

> The scenario started from the assumption that the climate negotiations in South Africa in 2011 had resulted in an international post-2012 agreement (in the form of a set of COP decisions) covering all parts of the Bali Action Plan including REDD. Now, in 2015, COP21 in Santiago de Chile was called to begin operationalising the 'REDD window' under a yet-to-be defined Global Green Climate Fund. Regarding REDD, the scenario reported that an adjusted historical baseline approach to reference level setting had been endorsed, as well as primarily national level monitoring and reporting. The specific setting of the policy exercise was a pre-meeting to the Santiago summit called by Brazil. The purpose of the meeting was to come up with a 'joint proposal for a REDD fund' that could subsequently guide the negotiations in Santiago de Chile.

Box 4. Summary of the scenario for the REDD exercise.

The role descriptions are summarised in Box 5 whereas the full versions can be found in Annex 1.H. The decision-making mode in the REDD-ALERT exercise was consensus-based, in line with what would be customary in this kind of international setting. This,

along with the fact that the exercise was framed as a preparatory meeting and not an actual negotiation, had an impact on the results of the run, as we will see below.

US in 2015: has re-engaged with the international climate regime, more activity on climate policy also domestically. Keen to see progress on REDD, but strictly performance-based, under a trusted international institution, and mainly relying on funding from carbon markets and the private sector.

EU in 2015: has stagnated politically and economically, climate change provides a key EU narrative. Positive climate policy record domestically, yet dismal in disbursing international climate funding. Wants REDD to develop into a robust and efficient mechanism, equipped with sufficient environmental and social safeguards.

India in 2015: has evolved into a G77 leader of sorts in the climate negotiations, also more proactive climate policy domestically. Stands to gain little from REDD, given its limited forest cover. Keen on preventing REDD credits from crowding out other international offsets, but recognises its importance as a trendsetter for an equitable governance structure of the Global Green Climate Fund (GGCF).

Brazil in 2015: strong performance over the past 5 years, economically and politically, inside and outside of the climate regime. Continues its pioneering role on REDD and has achieved a drastic reduction in its deforestation rate. Opposed to large-scale use of REDD credits for compliance, trusts in a global, performance-based approach.

Box 5. Summary of the role descriptions for the REDD exercise.

4.4.2 Runs and outcomes of the REDD-ALERT exercise

Run with experts

The REDD-ALERT policy exercise took place on 17 November 2010 in Amsterdam, involving 17 international participants. Country teams spent the morning defining their national position on the design of a global REDD+ fund. After an initial phase in which the teams worked individually to familiarise themselves with the material and to establish their country´s negotiating priorities, Brazil, as the host of the meeting, was the first to reach out to other Parties, starting a series of bilateral consultations. At mid-day, Brazil presented a ´non-paper´ which contained its proposal for a global REDD+ fund (see Table 8) and which was to form the basis for plenary discussions in the afternoon. The Brazilian proposal represented an attempt at striking a balance between satisfying developed countries' emphasis on a strictly performance-based fund on the one hand, and the interest of developing countries in securing sufficient levels of financing and in safeguarding the sovereignty over their resources and national development on the other. The former is reflected in relatively strict modalities proposed for verification (to be conducted by independent entities, rather

than based on countries' own reporting), the latter for instance in the determination of the payment size for achieved REDD reductions.

	Proposal by Brazil
Fund capitalisation and replenishment: *Where should the money for the fund come from?*	Revenues from auctioning emission rights in national emissions trading schemes as well as climate change bonds issued by developing country governments
Accounting for co-benefits: *Should high(er) social/environmental benefits of REDD activities be rewarded, and if so, how?*	All REDD units must be certified for high co-benefits by an independent third-party verifier
Eligibility criteria: *How to prioritise specific funding needs in view of overall constrained funding?*	Dedicated funding windows for different stages of REDD+ readiness, accounting for specific situations.
Fund administration: *Who should manage the fund?*	Global Environment Facility (GEF)
Verification: *Who should verify that REDD activities are realised and result in actual emission reductions?*	Accredited independent entities, similar to Designated Operational Entities under the Clean Development Mechanism.
Payment size: *How to determine the level of payment to REDD countries?*	Based on countries' own estimate of implementation costs, subject to review by an international panel, based on internationally accepted criteria.
Modalities for upfront financing: *Should REDD countries have access to upfront financing and if so, how should this be organised?*	Upfront payments to be issued as loans, transformed into grants if targets set out in national REDD+ strategies are met and verified.
Optional: Recommendation on the ***Transition of REDD into the global carbon market***	REDD+ fund acts as an 'investment reserve' (Obersteiner et al., 2009), purchasing yet-to-be-created REDD+ units upfront, then banking MRV-based credits until the fund's governing body decides that credits can be sold to the carbon market.

Table 8. Proposal by Brazil for a global REDD fund.

The ensuing plenary negotiations exposed the fundamental differences between parties on the basic rationale and functioning of a REDD fund. The United States on the one hand and Cameroon, often supported by India, on the other, occupied the two most extreme positions. The former was keen to see the fund operating as efficiently as possible, and to channel investments where they were most likely to

yield a return in terms of verified emission reductions from deforestation. The latter underscored the need for a comprehensive REDD system which would allow less developed countries to ´catch up´ on REDD-readiness and benefit equally from the mechanism. Brazil sought to move the process forward by separating issues on which consensus seemed in reach and others which were more deeply divisive. However, as time got short in the afternoon, parties realised that they would not be able to achieve consensus on all issues up for decision. Brazil as the chair of the meeting thus resorted to pushing for agreement on principles where more far-reaching consensus remained elusive.

Outcomes of the expert run

Table 9 summarises the 'joint vision for a global REDD fund' agreed by all teams. Owing to the consensus requirement, but arguably also to the style of chairing of Brazil, the vision is quite vague in many aspects. Even for design features where all but one country agreed, the final wording was watered down to receive unanimous backing of the whole group. Most concrete are the outcomes on verification of countries' performance under the REDD mechanism. This process is to be conducted by independent entities, similar to the arrangements under the CDM. The modalities for determining the payment size were also comparatively specific. Countries moreover endorsed a loan-grant mechanism for upfront payments; REDD countries could obtain upfront financing as loans which would be transformed into grants once the targets set out in the national REDD strategies had been met and verified. The question which body should administer the fund was fiercely debated over the course of the day; the final compromise endorsed the Global Environment Facility (GEF) as an interim financial mechanism, mirroring earlier decisions taken for other funds under the international climate regime. Countries also agreed on the long-term need for carbon market financing to keep REDD viable, yet without specifying how the transition from a fund to a market-based approach should be engineered. On the issue of eligibility criteria – who should be able to benefit from the fund –, the teams could only agree that it should take account of countries' different stages of development and REDD-readiness. Not surprisingly, it was the issue of fund capitalisation and replenishment where countries' positions were farthest apart. Agreement on specific sources of financing or commitments for certain country groups remained elusive; countries merely settled for the vague formula that finance for the global REDD fund should be 'stable, predictable, long-term, and sufficient'.

	Joint vision for a global REDD fund
Fund capitalisation and replenishment	Principles of stable, predictable, long-term, and sufficient funding.
Eligibility criteria	Taking account of different stages of development and

	different stages of REDD-readiness.
Fund administration	Criteria: efficient and credible/accountable to all parties. Global Environment Facility (GEF) proposed as interim mechanism.
Verification	To be operationalised as a priority. Carried out by accredited independent entities, who are accountable to the UNFCCC
Determination of payment size	Based on countries' own estimate of implementation costs, subject to review by an international panel, based on internationally accepted criteria.
Modalities for upfront financing	Loan-grant mechanism, taking account of LDCs.
Transition of REDD into the global carbon market	Carbon market financing needed to keep REDD viable in the long term.

Table 9. Joint vision for a global REDD fund agreed by all teams.

Country teams´ own evaluations to what extent this outcome met the success criteria that they had defined for themselves at the beginning of the day did not reveal great enthusiasm about the results reached.[21] India and Cameroon were slightly more positive than the EU, Brazil and US, yet compared to the average evaluations in the ADAM exercise (where both sub- groups evaluated their results with an average of 4.0 out of 5.0), the overall average of 2.98 out of 5.0 for the REDD-ALERT exercise seems quite low.

Nonetheless, based on the debriefing, participants found the results of the simulation to be plausible and they stressed that also in reality, agreement on principles would likely precede the operationalisation of a global REDD+ fund. Interestingly, in contrast to the other expert group, participants in the REDD exercise were positive without reservation about the policy element cards as a structuring device. As one participant commented during the debriefing, 'the cards helped bring some order into my thinking and our discussions on this issue which is still somewhat far from day-to-day business in the [UN climate, C.H.] negotiations.' On the whole, the policy exercise method was felt to have worked well to enable lively exchange and substantive discussions, although some participants expressed doubts about the extent to which they had succeeded in projecting themselves to the year 2015, as stipulated in the exercise scenario.

[21] Immediately after the exercise, country teams had been asked to revisit the sheet on which they had selected the five key criteria defining their country's success in the upcoming negotiations at the beginning of the day, and to rate on a scale from 1 to 5 to what extent the outcome reached satisfied each of the criteria.

The student runs

The student runs of the REDD exercise were held in November 2010, with a total of four groups going through the exercise on two different days. Given the technical complexity of REDD and the fact that this topic had not yet been treated in the master course that the students were following, we decided to somewhat narrow down the scope of the negotiations by reducing the design features to be discussed from eight to six. Hence, the questions of the 'determination of payment size' and the 'modalities for upfront financing' were omitted from the exercise. This proved a sensible decision as most country teams needed considerable time in the morning session to comprehend the basic concepts underlying REDD and their implications for their country´s position, and would have been hard pressed to analyse additional issues and options.

On the whole, the student groups went to greater length in trying to negotiate a concrete outcome than the expert group. Three out of four had agreed on a joint vision at the end of the afternoon session, while some features remained pending in the fourth group. Like in the expert run, the questions of how to capitalise the fund and whether and how to account for social and environmental co-benefits of REDD received most attention in the student simulations. The student groups negotiated very different outcomes, depending primarily on country teams´ interpretations of their roles, as well as their persuasiveness and perseverance during the negotiations. For instance, in one group, the US team strayed somewhat from its role description by displaying an exceptionally green-minded, multilateral orientation. In this case, the US eventually gave in to pressure from other parties and committed to mandatory contributions to the fund on a GDP per capita basis, a position far removed from the typical US stance in global climate politics. As in the case of the ADAM exercise, the substantive results of the student runs thus did not reflect fully realistic negotiation dynamics.

4.5 Summary

The two policy exercises we organised covered two crucial topics in the debate over how best to mitigate climate change. The discussions on both burden-sharing and the efforts to reduce emissions from deforestation and forest degradation entail pivotal trade-offs between efficiency and equity considerations that lead to predictably opposing interests at both the EU and global levels. During the design phase, we ensured that the set-up of the two respective policy exercises was largely similar, so as to minimise the effect that variation in the exercise design might have on learning outcomes. We subsequently ran these exercises with both expert and student groups. Whereas individual expertise and negotiation skills played a considerable role among both experts and students, negotiation dynamics were more realistic in the expert groups, leading to substantively more relevant (and plausible) results. This chapter

therefore focused on summarising the results from the expert runs. That does not mean that the student runs were not interesting however. As the subsequent three chapters on the different learning effects will show, the learning potential of policy exercises might actually be greater for students.

5. Cognitive learning

5.1 Introduction

Stimulating cognitive learning, among participants but if possible also in the wider policy community, is a core objective when designing and running policy exercises. This chapter describes the findings for this first type of learning effects in our typology. Cognitive learning was defined in Chapter 2 as the acquisition of new knowledge, the better structuring of existing knowledge or enhanced ability at systems thinking. In the context of this thesis, 'new knowledge' is considered to encompass insights into relevant facts, causal relationships or dynamics that are new to an individual participant of the policy exercise. Writings on simulation gaming and policy exercises claim that policy exercises also have the potential to produce knowledge that is 'new to the world' – whether this concerns the development of novel policy options, innovations to the policy process, or previously ignored cause-effect relationships (Geurts et al., 2007; Parson, 1996b; Underwood & Duke, 1987). This latter aspect not only represents a very high yardstick, but is also difficult to assess empirically; we reflect on this question for our own policy exercises in Section 5.3 below.

Integrating knowledge from different sources, developing a holistic view of a problem in all its dimensions, quickly exchanging and verifying facts, establishing priorities for action, developing a long-term perspective – the list of alleged cognitive learning benefits of policy games and exercises in the literature is long (see also Chapter 2). Often-cited reasons for their ability to stimulate cognitive learning are the time pressure that participants experience during a policy exercise and the concrete tasks they need to resolve. This 'requirement to act' arguably leads participants to 'use available sources of knowledge more aggressively, integrate them more broadly, and to organize them more coherently to support their required decision' than in conventional workshop settings (Parson, 1996b, 244). By experimenting with different policy options, participants can gain knowledge about how these options are perceived by other stakeholders, what criteria might be applied to evaluate them as well as their possible effects (Brewer, 1986; Parson, 1996b). Some authors also stress the potential of these methods to uncover the 'hidden face in the picture' (Jones, 1964; Schelling, 1964), for instance by revealing unintended consequences of a particular course of action or by drawing attention to previously ignored aspects of the problem (Brewer, 1986; Mayer, 2009). Finally, the scenario element inherent in policy exercises may enhance a long-term orientation in participants' thinking about a problem, especially if they walk forward in time through several rounds of scenario-building (Geurts et al., 2007).

Obviously, the type and extent of the cognitive learning benefits of a policy game depend on its specific objectives and hence its design. As the focus of our own policy exercises was on assessing different policy options and their feasibility for EU burden-sharing and REDD, the expectation was that insights on these matters would constitute a key area of participants' cognitive learning. Arguably, this would extend both to current policy – since the discussions during the exercise would necessarily reflect back on the effects and implications of present policy design – as well as future developments, given the future setting of the policy exercise. Moreover, we expected to find learning about policy dynamics especially in the student groups that were less familiar with the policy-making processes (EU decision-making and global negotiations) simulated in the two exercises. Finally, this chapter aims to generate insights into the 'knowledge structuring' dimension of cognitive learning. As Chapter 2 argued, cognitive learning involves not just the assimilation of new factual knowledge, but also the improved ability to structure existing knowledge. This may manifest itself on the one hand through a more differentiated mental map, i.e. through more associations and more causal connections in relation to a specific aspect. On the other hand, improved knowledge structuring may also involve better being able to synthesise information, i.e. making sound decisions on what may or may not be relevant in a particular context.

In terms of differences between the four groups, in line with the propositions on expected learning outcomes presented in Chapter 3, it seemed logical to expect higher levels of cognitive learning from the students compared to the expert groups, given their lower level of prior expertise. The impact of a one-day intervention on the expert participants, all of whom with years of experience on the subject, would conceivably be more limited. As a second proposition, it was assumed that the difference in terms of topic – EU burden-sharing vs. REDD financing – would, if at all, matter only for the expert groups. This is because, whereas both topics were equally unfamiliar to the students, REDD financing as the more unstructured topic that leaves more space for 'blue sky' solutions and creative thinking (see Chapter3) might offer greater potential for cognitive learning in the expert group.

In analysing cognitive learning, we drew on a number of different data sources. Comparison of concept maps drawn prior to and after the exercise allowed for insights in participants' conceptual change as a consequence of the exercise as well as on knowledge structuring. Furthermore, we make use of self-reported information on cognitive learning from post-questionnaires and interviews with participants, drawing on our own systematic observations during the workshops in order to explain the patterns that we find. The structure of this chapter reflects this. We begin by introducing the notion of concept maps and different ways of using and analysing them. We then describe the way in which we have employed them in our research and subsequently report on the outcomes for the four participant groups. Subsequently, we move on to discussing our findings on participants' self-reported

cognitive learning, based on survey and interview data. Finally, we reflect on the relevance of various contextual factors for the levels of cognitive learning that we found.

5.2 Tracing recorded cognitive change using concept maps

5.2.1 Concept maps as a knowledge structuring tool

Concept maps are a graphical device for representing knowledge on a certain content domain and in this sense related to other, similar methods like mind maps and cognitive maps.[22] A concept map typically comprises nodes representing concepts and linking lines which stand for relationships between these concepts. Concept mapping is a popular tool not only in education, but also in business and governmental contexts, and has been used for a variety of purposes, from instruction and assessment to knowledge elicitation and capture, to group brainstorming (Cañas et al., 2003).

The notion of concept maps goes back to work by Novak and Gowin (Novak & Gowin, 1984; Novak, 1998), who developed them as a methodological extension of Ausubel's work on learning psychology (Ausubel, 1962). This theory assumes that learning occurs as a person integrates new propositions and concepts into his or her existing normative and conceptual framework, thus enhancing his or her cognitive structure. Concept maps make some of the main elements of this assimilation process visible: first, there is the process of subsumption, which refers to the semi-hierarchical structure of knowledge, where broader concepts subsume more specific ones, thus serving as an ordering device. A second element, integrative reconciliation, implies the fitting of new propositions into the existing framework, with the latter being modified where necessary. Finally, progressive differentiation describes how new concepts add to the complexity of the cognitive framework by creating new branches, or adding more levels of hierarchy. In an assessment context, concept maps make these processes visible as they 'tap into a learner's cognitive structure and to externalize, for both the learner and the teacher to see, what the learner already knows' (Novak & Gowin, 1984, 40). This structural aspect of knowledge, which assumes that having significant expertise in a domain implies knowledge that is highly

[22] The distinction between mind maps, concept maps, and cognitive maps is fluid, and usage of the terms differs somewhat between authors. Mind maps, initially conceived by Buzan and Buzan (1993), are tree-shaped constructs departing radially from a central idea. They are primarily used for stimulating creativity in brainstorming or for planning purposes. Cognitive mapping was first promoted by Colin Eden and his colleagues (Colin Eden, 2004; Rohrbaugh & Eden, 1990) in an operations research context and is often referred to as the process of developing shared mental models in groups.

integrated and structured around central concepts, is widely recognised in cognitive learning theory (e.g. Glaser & Bassok, 1989; Ruiz-Primo & Shavelson, 1996).

Concept maps as an assessment tool have so far been almost exclusively used in an education context, ranging from high school students (Kinchin et al., 2000; Rice et al., 1998; Yin et al., 2005) to university courses ranging from educational psychology to biology and nursing (All & Havens, 1997; Derbentseva et al., 2007; McClture et al., 1999). To our knowledge, the only earlier application inn a context similar to ours was by Huitema et al. (2010), who used pre- and post-concept mapping to assess participants' learning in a series of citizen juries. Their analysis focused more on changes in map structure, however, paying less attention to shifts in content.

The basic logic of concept mapping as an assessment device is the same across most studies – participants are asked to complete a concept map prior to an intervention, and a second one afterwards. Yet as the use of concept maps has proliferated over the past decades, the format of the concept maps and the methods to analyse the observed changes have become increasingly varied. While some maps are built up hierarchically with the central concept on top, others are organised radially around a core. Connecting lines may be labelled (this is called a propositional mapping format), or not. Finally, map completion can be a more or less 'directed' process: participants may be asked to fill in 'scaffold' maps with given concepts and linking terms, or, more commonly, draw a map 'from scratch' on a blank sheet of paper (Ruiz-Primo & Shavelson, 1996). When it comes to map assessment, one group of methods, building on Novak and Gowin's initial work, scores maps based on the number of components and their structure, distinguishing between valid and invalid propositions and crosslinks, either by counting only those deemed accurate or by rewarding them with a higher score (Pearsall et al., 1996). A second cluster of methods assesses participants' maps based on their similarity to a 'master map', through manual or computerised comparison with the concepts and propositions, or the proximity of certain concepts, on the latter (Acton et al., 1994; Luckie et al., 2011). Yet others seek to combine both approaches and score components, but also use an expert map to weigh them (Rye & Rubba, 2002). Finally, some authors rely on a global assessment of the map by one or more experts, or evaluate maps based on certain general criteria (e.g. Lomask et al., 1992).

5.2.2 Approach to concept map analysis in order to assess knowledge structuring

In devising our approach to using concept mapping to assess cognitive learning, we had to account for the specific objectives and context of our research. The primary goal of our policy exercises was to advance knowledge on two specific aspects of European and international climate policy, burden-sharing and REDD financing. We sought to accomplish this through a flexibly structured format which emphasised interaction and exchange among the participants. The exercises were not an

instruction setting aimed at conveying pre-specified sets of information to participants, as concept maps have traditionally been used. This had implications both for how we presented the task to participants and how we went about analysing the resulting maps. A less directed 'mapping from scratch' process, which had participants create their own map on a blank sheet of paper, seemed more appropriate to the nature and objective of our intervention. In the same vein, to keep the exercise 'light' and as little 'test-like' as possible, we opted for a pure mapping rather than a propositional mapping format, which did not require participants to label the connecting lines on their maps. Since we were interested in *what* participants learned rather than ascertaining whether they learned 'the right things', we felt that scoring the maps based on a comparison with a 'master map' drawn up by the organising team would not be the right approach. Rather, we searched for an established methodology that would allow for an aggregated analysis of the content and structure of the maps, without judging the perceived accuracy of the connections made.

These requirements led us to opt for a methodology that had initially been developed by Morine-Dershimer (1993) to trace conceptual change in teacher education, complementing it with some additional metrics on map structure. The method relies on a coding of the items on the sets of pre-and post- concepts maps and is based on two key principles, (1) centrality – the proximity of certain concepts to the core of the map, which can be taken as an indicator how salient they are in the perception of the author and (2) specificity – the proportional frequency with which these concepts occur on the map. A comparison of shifts in the centrality and specificity of concepts from the pre- to the post-measurements therefore allows tracing changes in the structuring of knowledge as well as newly perceived relevance of certain aspects. The latter is an important aspect of our definition of 'new knowledge', which is not limited to the appearance of entirely novel concepts on participants' mental maps, but which can also manifest itself through increased awareness of the salience of particular aspects of the problem.

Following Morine-Dershimer, we started out by coding all items on the concept maps into a number of categories which capture the content on the maps. The categories were developed inductively, by working iteratively and adjusting the categories and their definitions until all items had been accommodated in a consistent and coherent manner. Given the different topics of the two policy exercises, we devised two separate coding systems, both of which ended up having eleven categories. An explanation of the two coding systems, including sample responses for each category, can be found in Annex 2.A.

Subsequently, for each category on every map, scores for centrality and degree of specificity were calculated. Centrality scores were calculated based on the level at which the category first appeared on the map, relative to the map core. For example,

if a concept was linked directly to the core, its centrality score was 1. If another concept first appeared in a reference connected to the first concept, its centrality score was 2, and so on. In order to enable comparison of scores across concept maps and participants, we needed a measure for each concept category from each participant, even if that category did not appear on a map. Therefore, again following Morine-Dershimer, concept categories that were absent on a map were counted as being two levels below the furthest level present on that map. For example, for a map that extended to five levels, a category that had no items noted received a centrality score of seven. Specificity scores were calculated based on the relative frequency of items associated with one category – the number of items falling under a specific category was divided by the total number of items on the map. For instance, for a map with a total of 30 items, six of which were coded as belonging to the category 'international context', the degree of specificity for this category on this map was .20 (twenty per cent of items on this map belonged to this category). Along the same lines, if a category did not appear at all on a map, it received a specificity score of .00.

Participants were asked to draw one concept map before the policy exercise and a second one afterwards. During the workshop dinner on the eve of the workshop (at the morning of the policy exercise for the student groups), participants spent ten to fifteen minutes constructing a concept map on the topic of the policy exercise. They were handed a sheet of paper, the front side of which was empty except for a short explanation of concept mapping and a central node containing the topic of the policy exercise ('Key aspects of EU burden-sharing' and 'Setting up a global REDD fund', respectively), that was to form the point of departure for a radially arranged map. The back side of the sheets contained an example of a concept map unrelated to the topic of the exercise to help participants unfamiliar with concept mapping to get started. After the policy exercise, participants were asked to draw a second concept map, with the same instructions as at the beginning of the day. In the expert run of the burden-sharing exercise, three participants who had to leave to catch their trains or planes home did not draw the map on the spot, but sent scanned maps via email a few days after the exercise. Similarly, since the students in both exercises had completed their first map only in the morning, they were told they could draw their post-maps at home and submit them during the following week. Neither experts nor students had their first map available when drawing the second one.

With sets of 60 and 69 paired pre- and post-concept maps for the EU Burden-sharing and the REDD policy exercise, respectively, the response rate was quite high for the student groups, as for them this was a compulsory part of the course they were following. For the expert groups, while the response rate for the ex ante maps was close to 100%, the rate afterwards was considerably lower, owing to the difficulty of motivating participants for yet another task after a long and strenuous workshop day. In the end, we obtained sets of comparable pre- and post-maps from five participants

for the EU burden-sharing exercise, and nine pre- and post-sets from participants for the REDD exercise. This constitutes a major limitation to the validity of our findings for these groups and we discuss the implications of this below. We will now turn to describe the results of the concept map analysis, first for the runs of the EU burden-sharing exercise and then for the REDD exercise, before making some observations across all four groups.

5.2.3 Cognitive change in the policy exercises on EU burden-sharing

Run with experts

As stated above, with only five available sets of pre- and post-concept maps, the response rate for the expert run of the EU burden-sharing exercise was very low. However, we obtained at least one set from each occupational group present at the policy exercise (policy-makers, academics, consultants, private sector), which enhances the representativeness of this small sample to some extent. Figure 5 displays mean centrality and specificity scores at pre- and post-test for the available maps (n=5) (for a table including all scores, see Annex 2.B, Table 23). Mean centrality scores are depicted on the x-axis, with centrality decreasing from left to right. Mean specificity scores are indicated on the y-axis – the higher a category on the map, the more frequently it is referred to on average across the available maps. Thus, categories that moved left have become more central at post-test, while those that have moved right have become less so. Along the same lines, categories that moved up have increased their proportional frequency across maps, whereas those moving down appear less often on ex post maps.

Looking at Figure 5 below, we can observe that a number of concept categories shifted from pre- to post-measurement for the expert group. Simultaneous shifts on both indicators, centrality and specificity, were most noticeable for the two categories relating to the design and implementation of burden-sharing arrangements and to the European emissions trading system (EU ETS).[23] Both

[23] The results presented here differ from the ones reported in Haug et al. (2011). For this paper, we had coded the map items and analysed the resulting dataset from the expert run only, before even collecting the data from the student run on the same subject. Yet when attempting later to apply this initial coding scheme to the student maps, we realised that it did not fully capture the content contained in the student maps. We thus adjusted and expanded the coding system in order to account for the diversity of items across the whole set of maps. In doings so, we inevitably lost some of the subtleties and nuances in the results for the expert group – thus arguably underreporting the cognitive change among this set of participants –, yet felt that the benefits of achieving comparability in the results from both groups outweighed this disadvantage. For the concept map data from the REDD exercise, we worked from the outset with maps from experts and students to develop common coding categories applicable to both participant groups.

became more central as well as specific on the ex post maps. This resonates with our observations during the policy exercise, where discussions circled extensively around the question how to design a burden-sharing agreement that is robust (containing sufficiently stringent enforcement mechanisms) yet flexible enough to accommodate for unforeseen developments (e.g. particularly cold winters or hot summers, which might result in emissions increases through increased heating/air-conditioning).

Similarly, over the course of the workshop, participants became increasingly aware of the strong interrelation between burden-sharing and the functioning of the EU ETS, which again explains why these issues became more central and specific on the ex post maps. On the other hand, the specificity of the treatment of criteria for burden-sharing, while still very high on the post-maps, decreased considerably. Items relating to this category made up more than thirty per cent of all concepts on experts' ex ante concept maps. On the ex post maps, their (still sizeable) share decreased to slightly above 20 per cent. Looking at the maps in more (qualitative) detail, we find that pre-maps were dominated by rather generic lists of possible criteria (cost-efficiency, cost-effectiveness, equity…) on which future burden-sharing schemes could be based. These lists were replaced by a more diverse outlook on the topic in the post measurement, as is noticeable from the spread of concept categories in lower part of Figure 5. In section 5.2.5, we delve in more detail into the type of items contained in the maps (making a distinction between lists, labels, and explications) on the maps and discuss how shifts between types relate to the knowledge structuring property of cognitive learning.

In line with the approach introduced by Morine-Dershimer (1993), we subsequently sought to statistically assess the shifts in centrality and specificity using analysis of variances (ANOVA), with time of mapping (pre or post) as the independent variable. The analysis reveals differences between pre- and post-measurement in the degree of specificity for the EU ETS and the policy design and implementation categories, which are significant at $p <$. 05 level for EU emissions trading and at $p <$. 10 for policy design and implementation (see Annex 2.B, Table 24). It is likely that the small number of available concept maps prevented other, smaller effects from becoming statistically significant.

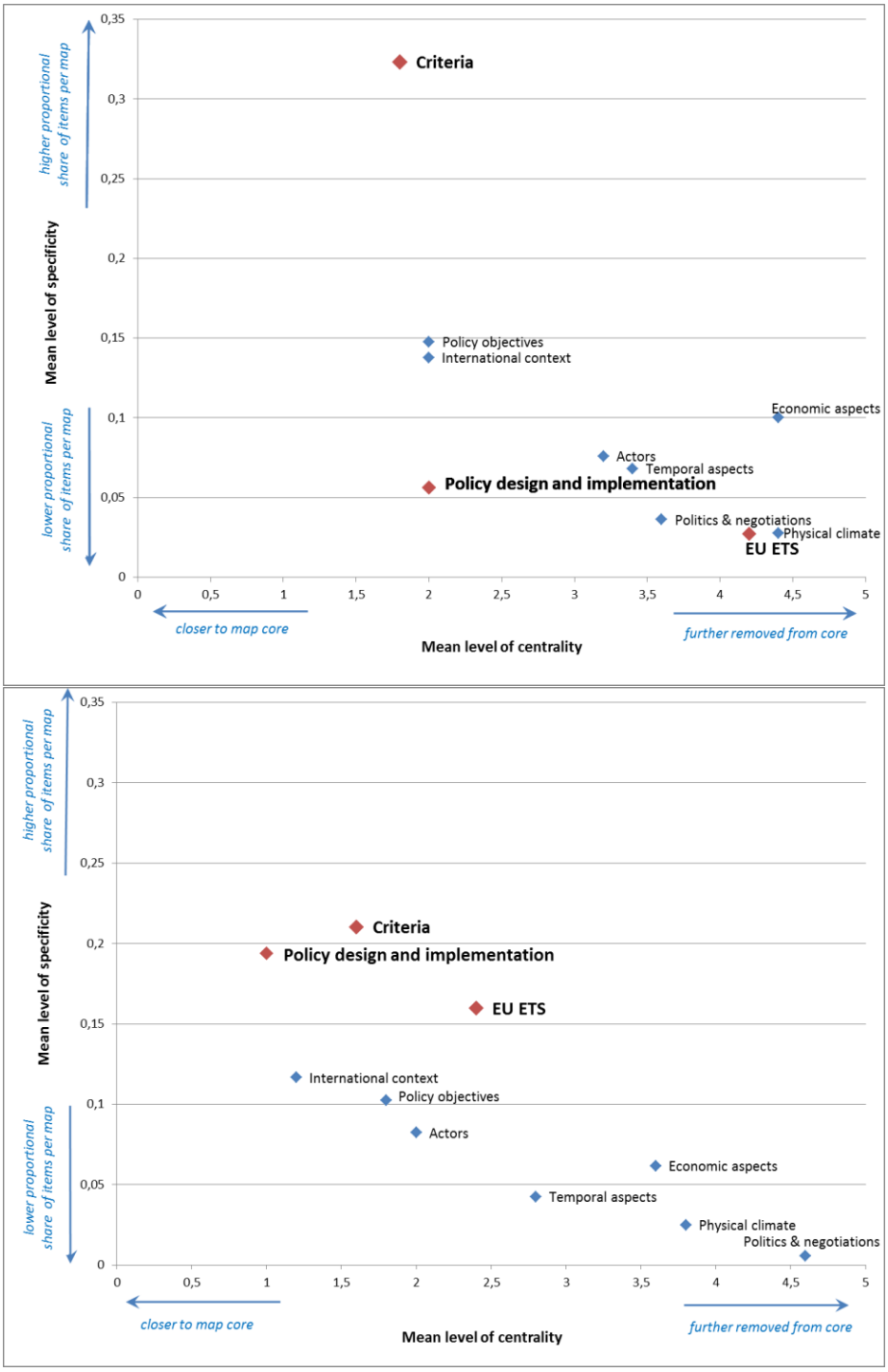

Figure 5. Centrality and specifity scores at pre-test (above) and post-test (below), experts burden-sharing exercise. Conceptual shifts discussed in the text are bolded.

Run with students

Turning to the concept maps drawn by the students who participated in the policy exercise on EU burden-sharing, we find somewhat different patterns (Figure 6; Table 25 in Annex 2.B). The most visible change is probably the rise of the 'politics and negotiations' concept category, which is statistically significant at $p <. 05$ in terms of specificity and at $p <. 10$ in terms of centrality (see Annex 2.B, Table 26 for all ANOVA results of the student group). This emphasis on the aspects related to the dynamics of politics and negotiations resonates with our findings on students' self-reported cognitive learning from the policy exercise, on which we report in more detail later in this chapter. The main similarity in concept mapping outcomes between the expert and the student group lies in the increased priority attached to aspects related to the EU emissions trading scheme. As for the experts, this aspect has become significantly more central and specific on the ex-post maps, reflecting the focus of much discussion on the interrelation between burden-sharing and the emissions trading system.

Two further statistically significant changes relate to the decrease in a proportionally smaller number of items (i.e. decreased specificity) for the categories 'actors' and 'policy objectives' on the ex post maps. When qualitatively examining the content of these categories, we note again a reduction in the 'listing effect' from pre- to most-measurement. Many ex ante concept maps had included extensive lists of actors ('industries', 'rich countries', 'poor countries', 'NGOs', 'governments', etc.) as well as a rather large number of highly generic references like 'climate target EU', 'goal-setting', or 'common aims', which we subsumed under the category of policy objectives. These lists, which in some cases seemed to serve the purpose of 'filling up' a rather empty ex ante map, had become more condensed and reduced to the essential on the post maps. The conceptual picture at post-test thus captures a more nuanced, substantive understanding of the issue.

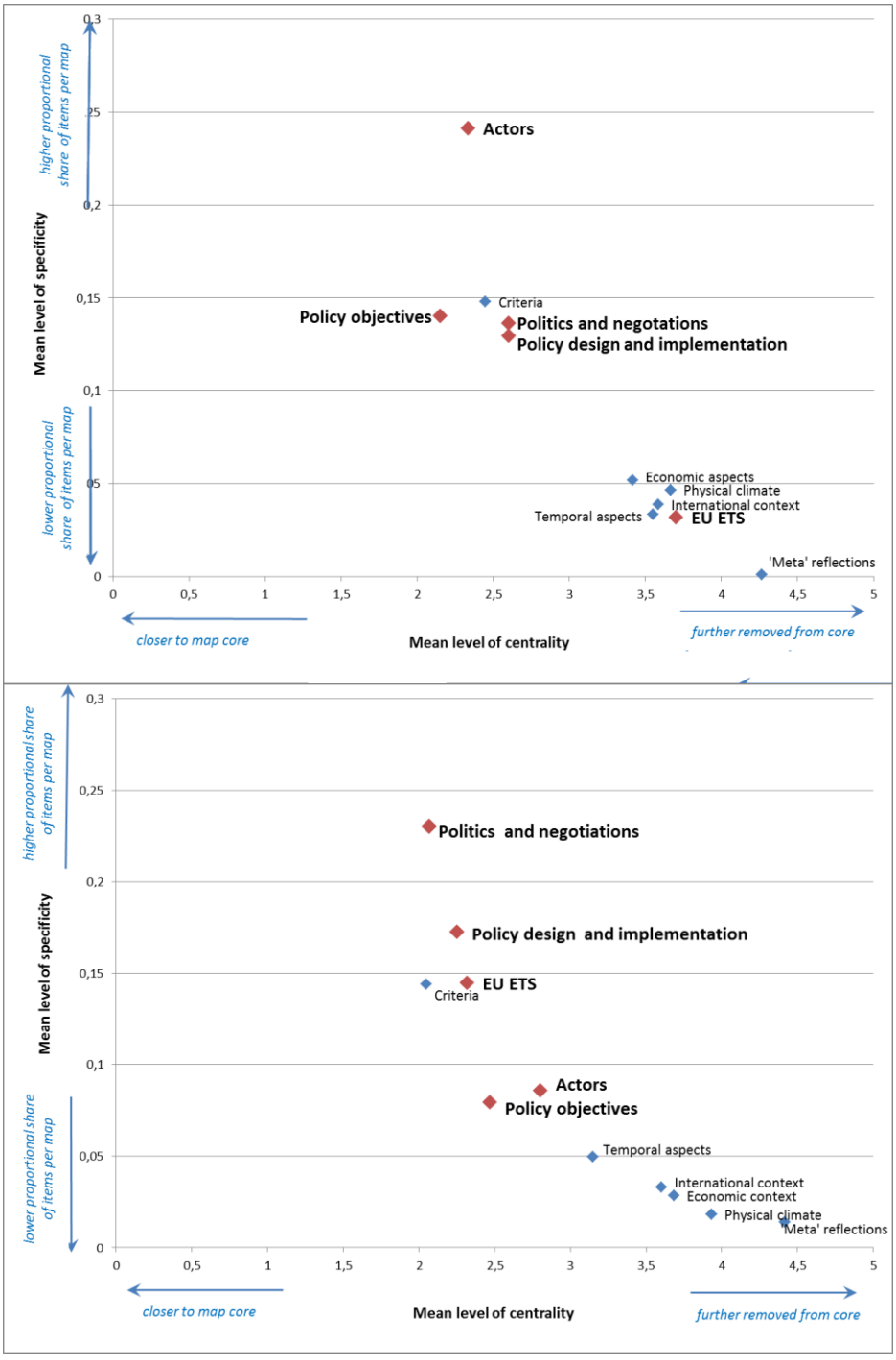

Figure 6. Centrality and specifity scores from pre- to post-test, students burden-sharing exercise. Conceptual shifts discussed in the text are bolded.

Having analysed the conceptual changes of expert and student group of the EU burden-sharing exercise separately, we were interested in comparing the patterns between the two groups. Independent-measures analysis of variance (ANOVA), using group affiliation (expert/student) as the independent variable, revealed a number of differences (see Annex 2.B, Tables 27 and 28). First of all, the category 'international context' was significantly more prominent on the expert maps at both pre- and post-test. We think that this is due to the richer contextual knowledge of this group that rightly attributed some importance to the global situation and policy developments, whereas the students tended to strongly focus on the task they were confronted with, namely the negotiation of an intra-EU agreement.[24] Moreover, the high priority the students attached to the politics and negotiations dimension is reflected in significantly higher specificity (at both pre- and both-test) and centrality scores (only at post-test). A final statistically significant difference relates to the much greater specificity which which the 'criteria' of and 'design and implementation' for burden-sharing agreements were elaborated on the expert pre-maps. This difference had all but disappeared in the post-measurement, however, testifying to a degree of convergence between expert and student knowledge following the exercise.

5.2.4 Cognitive change in the policy exercises on REDD

For the policy exercises on REDD, we followed the same approach to data collection as for the workshops on burden-sharing, both for the expert and the student runs. Since the topic of the exercise was different, the coding scheme for assessing the concept maps that we developed necessarily differed from the system for the burden-sharing exercise. An overview of the coding categories, including explanations and sample responses, can be found in Annex 2.A.

Run with experts

While still relatively low, the response rate, nine paired sets of pre- and post-maps, was somewhat better for the REDD policy exercise than for the one on burden-sharing. Still, the limited sample size may have prevented us from detecting medium or small statistical effects. The only significant shifts relate to the category 'actors'. The centrality of the concept decreased significantly from pre- to post map as did its specificity, although this was only at the $p < .10$ level. Again this is explained through a reduction of the 'listing effect' for this category; as participants learned in more depth about the topic over the course of the exercise, rather generic map items like

[24] The scenarios framing the policy exercises described the international policy context in which the exercise was set, and the journalist equally published news relating to international developments. However, as discussed in Chapter 4, these attempts to draw attention to the global dimension of the exercise seem to have had less impact on participants than we had anticipated.

'industrialised countries', 'developing countries', 'companies' and 'local stakeholders' diminished in number and centrality.

Figure 7 displays aggregated centrality and specificity scores for concept categories at pre- and post-test (a complete set of mean centrality and specificity scores is included in Annex 2.B, Table 29). We see that while the 'allocation and disbursement' category – subsuming items related to the expenditure side of REDD financing – has remained the most central and specific category in both pre- and the post-measurement, the supply side of REDD finance ('financing and sources') has moved up to second place on the post-maps. 'Governance of the mechanism' – relating to the institutional arrangements for managing REDD finance – also increased markedly in centrality and specificity. In contrast, there is one category, 'international context', which clearly lost in importance from pre- to post-measurement; especially specificity scores went down in this case. Taken together, these changes render visible the imprint that the discussions over the course of the policy exercise left on participants. They reflect the rather focused theme of the workshop, how to design a global REDD fund. Key questions in this regard – where should the funds come from, who should receive them and how should this process be organised – emerge as the main themes from the ex post concept maps. Important contextual factors, such as the international context for REDD, have lost in importance. A final category that gained on both indicators is 'policy objectives'. We are not sure how to interpret this, but the increase may be related to a theme in discussions throughout the workshop day on how to prioritise and make trade-offs the various objectives of REDD, from a focus on climate mitigation to a more integral vision that also accounts for biodiversity and livelihood goals.

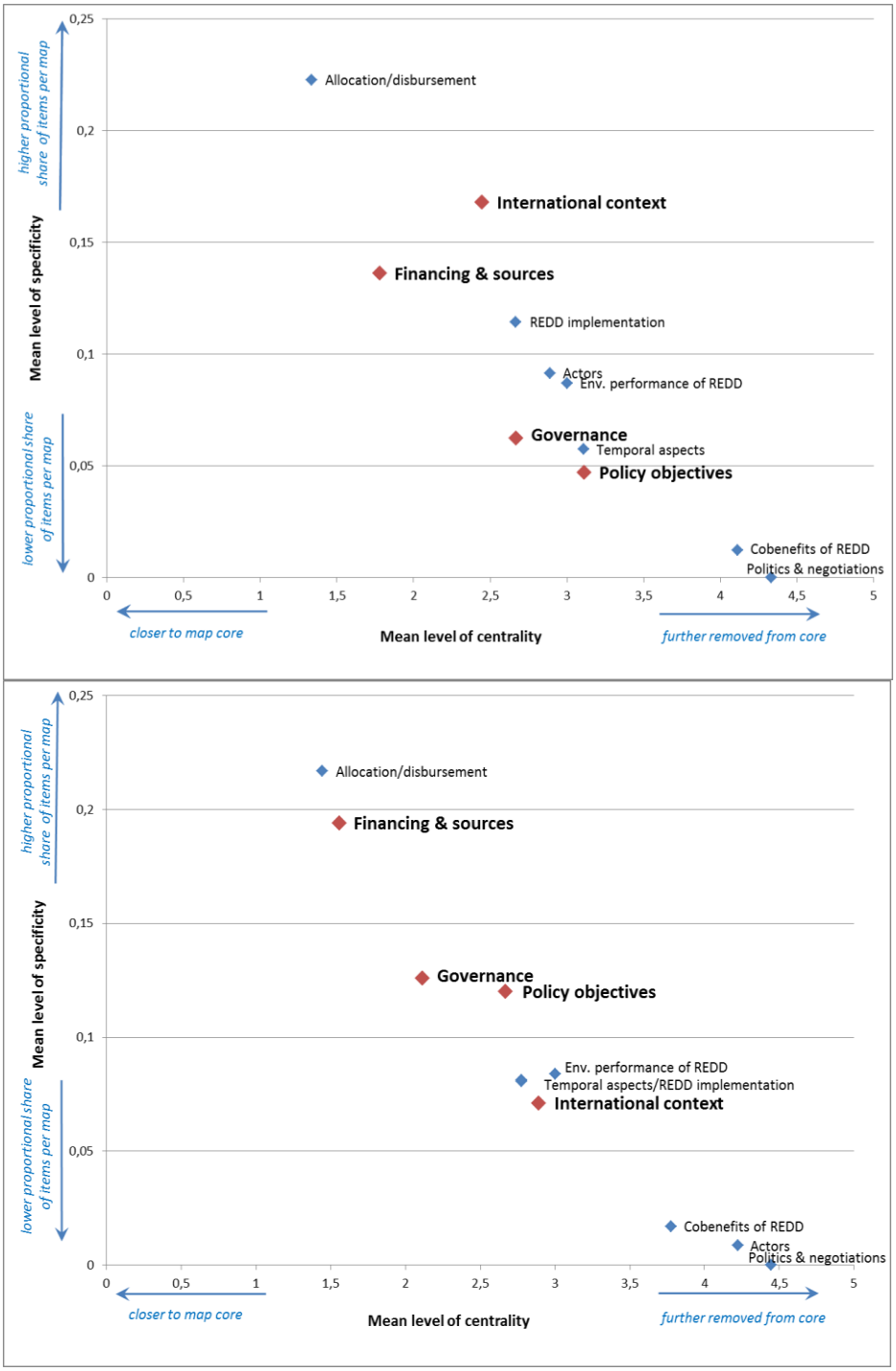

Figure 7. Centrality and specifity scores from pre- to post-test, experts REDD exercise. Conceptual shifts discussed are bolded.

Run with students

Turning to the concept maps drawn by the students who took part in the policy exercise on REDD, the visible shifts in centrality and specificity on Figure 8 do not seem particularly remarkable at first glance (see Annex 2.B, Table 31 for all scores). The two most central and specific categories – 'financing and sources' and 'environmental performance of REDD' – have by and large maintained their positions. However, analysis of variances pick up a rather large number of statistically significant changes in the dataset (n=69) (Annex 2.B, Table 32).

Two statistically significant upward shifts (in terms of both centrality and specificity) from pre- to post-measurement concern the categories 'governance of the mechanism' and 'co-benefits of REDD'. The increase in attention to these concept categories in the student group ties in with our findings on self-reported learning, which we will discuss later in this chapter. The increase in the perceived salience of the concept of co-benefits of REDD (relating to the question to what extent REDD mechanism design should reward the achievement of other social and environmental objectives beyond climate mitigation) also resonates with our findings on recorded normative learning (reported in the next chapter) and self-reported cognitive learning. By contrast, categories that decreased significantly on one or both measures include 'actors', 'REDD implementation', and 'policy objectives'. Taken together, we can interpret these findings along the same lines as the ones for the expert group. They reflect a strong instrumental concern with global REDD mechanism design, which stood central in the discussions over the course of the day. Against this background, more contextual categories ('actors', 'REDD implementation' – referring to the realisation of REDD projects on the ground) lost in prominence on the post-maps.

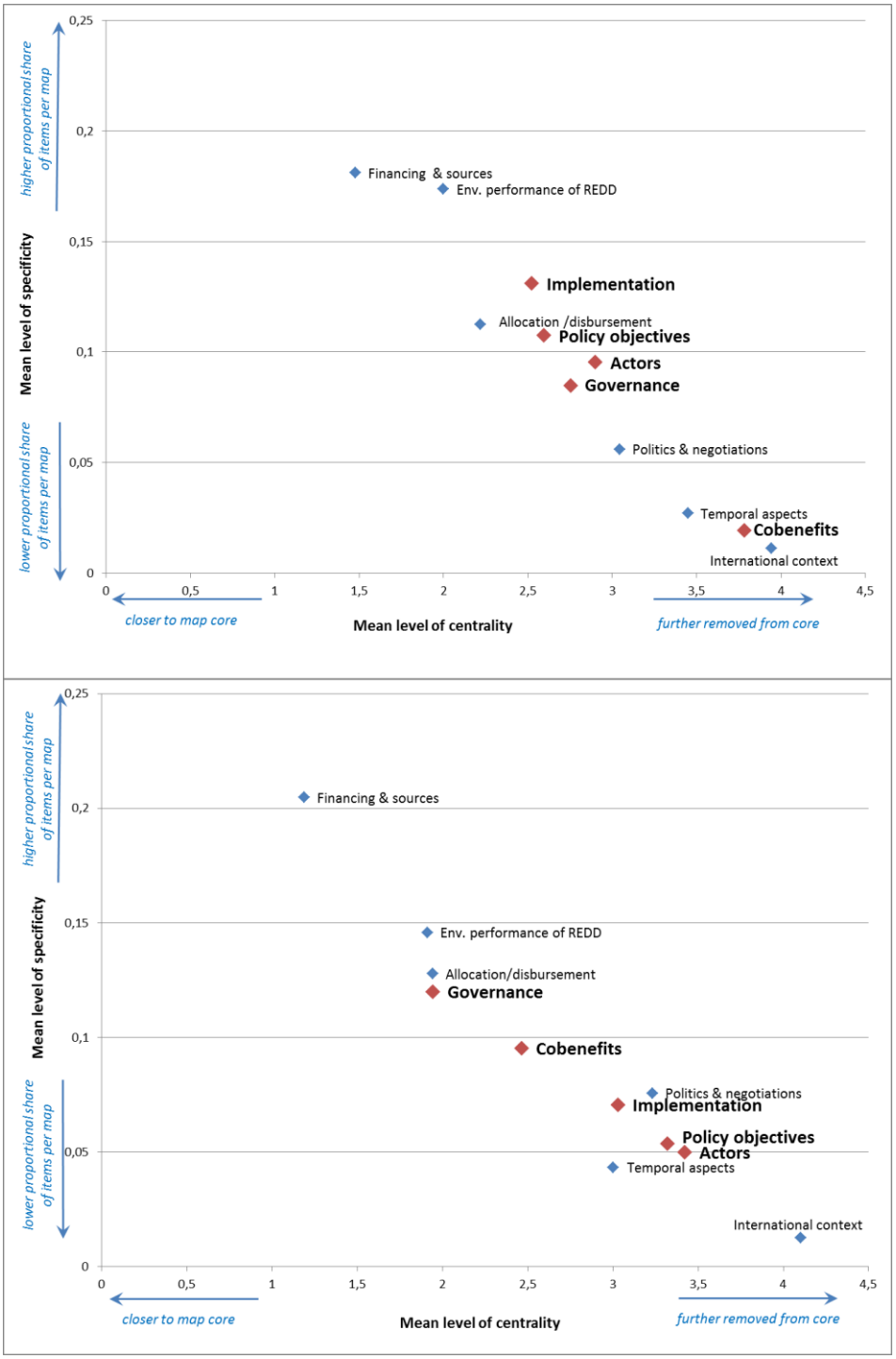

Figure 8. Centrality and specificity scores from pre- to post-test, students REDD exercise. Conceptual shifts discussed in the text are bolded.

Again, we were also interested in comparing the patterns of reference of experts and students to different concept categories. Overall, there is a relatively large number of statistically significant differences between the two groups, both at pre- and at post-test (see Annex 2.B, Tables 33 and 34). As for the burden-sharing exercise, the international context was significantly more prominent on expert maps, both for the pre- and the post-measurement (for the latter, only the difference in centrality was statistically significant). At pre-test, experts also included items related to 'allocation and disbursement' significantly more frequently on their maps than the students. This difference had diminished and was no longer statistically significant at post-test, however. By contrast, items that belong to the ´politics and negotiations´ category were significantly more central on students´ pre-maps, both before and after the exercise. The same applied for the categories 'co-benefits of REDD' and 'actors', but only at post-test. Overall, contrary to the burden-sharing exercise there is thus little indication of a convergence of experts' and students' mental maps as a consequence of the REDD exercise. One potential explanation for this relates to the nature of the two topics: a game on burden-sharing as the potentially more structured issue may hold more similar lessons for all participants, whereas with REDD financing, different aspects seem to have attracted the attention of experts and students, respectively.

5.2.5 Additional evidence for cognitive learning from the concept maps

As we have seen above, Morine-Dershimer's measures of centrality and specificity provide a valuable avenue for tracing the occurrence of conceptual change and also allows for insights regarding their content and direction. What we cannot deduce from the results of this method, however, is the extent to which these changes constitute an objective 'improvement', i.e. a 'better structuring of existing knowledge' compared to the status ex ante. As we argued in Chapter 2, we refrain from using a normative yardstick on whether participants learned 'the right things', adopting an open attitude regarding the question of 'what is learned'. Therefore, we decided against grading the quality of the concept maps or making use of a master map, which would have implied imposing the 'correct' content of a concept map or the 'right' way of structuring it. Instead, we searched for other indicators that might provide evidence for cognitive learning.

Number of items and levels per map

We first turned to comparing the total number of items and the number of levels (distance of nodes from the central node) on pre- and post-concept maps. Both metrics are used in a number of concept map assessment methods (e.g. Jacobs-Lawson & Hershey, 2002; Roehler et al., 1988). Although there is not a one-on-one relationship, we would posit that an increase in the number of levels speaks to an increased, and presumably improved, structuring of the drawer's mental map of an issue, thus attesting to progressive differentiation in the terms of Ausubel's learning theory (see Section 5.2.1 above). Similarly, more items on ex post concept maps may

be interpreted as evidence for a richer understanding of the topic at hand. This is not entirely unambiguous, however; as Huitema et al. (2010) rightly contend, cognitive learning may in some cases also be reflected in the elimination of false assumptions and irrelevant items.

Figures 9 and 10 present the number of items and levels as averages per participant group at pre- and post-test. We can see that the average number of items and levels on the concept maps increased from pre- to post-measurement in all four groups – with the exception of the mean number of levels for the expert group on burden-sharing, which remained constant (but at a higher level than for the other groups). For the two student groups, the increases from pre- to post-test (both in the number of levels and in the number of items) were statistically significant (see Annex 2.C, Table 35).

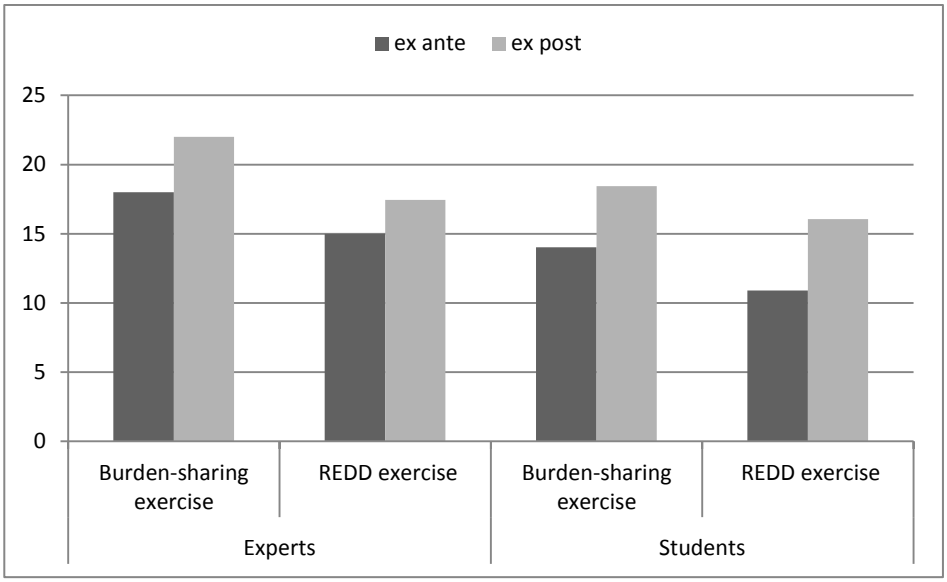

Figure 9. Average number of items per concept map at pre- and post-test.

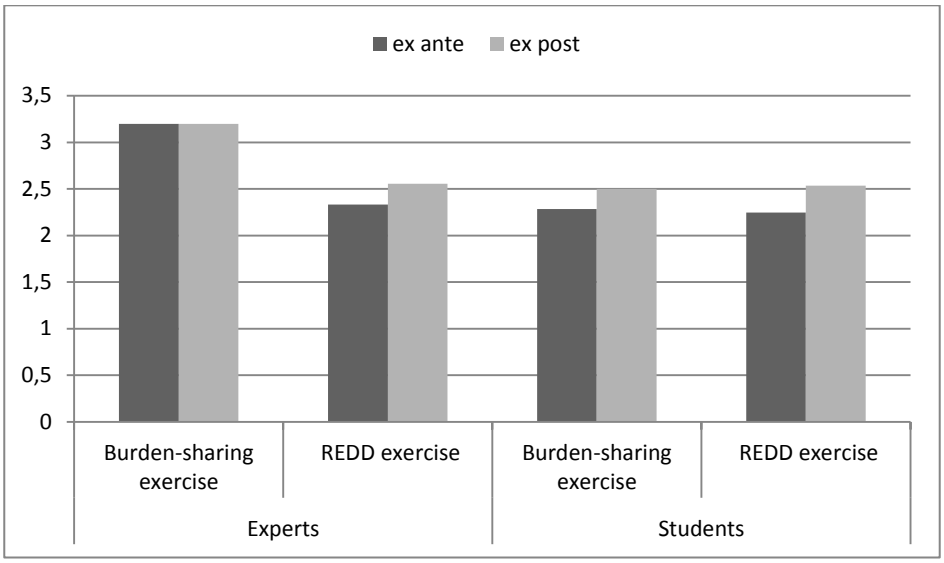

Figure 10. Average number of levels per concept map at pre- and post-test.

Can we detect a difference between experts and students on these indicators? We find that at pre-test, experts noted on average more items than students (16.07 vs. 12.57) and that their maps also had more levels (2.64 vs. 2.26), though both effects were only weakly significant at $p < .10$ (see Annex 2.C, Table 36). We cannot detect such effects at post-test any more, the gaps having narrowed both for the average number of items (19.07 for the experts vs. 17.33 for the students) and levels (2.79 vs. 2.52). We would interpret this as evidence that, in line with our expectations noted above, the policy exercise had a greater cognitive learning effect on the students than on the experts; the number of levels and items on the concept maps at post-test show an extent of convergence between experts' and students' knowledge on the topic.

Differentiating between different types of concept map items

In an additional step, to add a more qualitative element to the assessment of map quality, we used a distinction between different types of responses items that Morine-Dershimer introduced in her 1993 paper. She differentiates between three kinds of items on participants' maps: listings, labels and explications. Listings are extended enumerations under a superordinate item, which cite examples of this item, but hardly shed additional analytical light on it (e.g. 'housing, agriculture, electricity' under 'sectors' in the EU burden-sharing exercise). Labels in Morine-Dershimer's definition are generic, one-word statements (in our context for instance 'fairness', 'developing countries', 'negotiations'), without additional context or information. Explications refer to items including more elaborate comments, or clusters of items

that describe a higher-order item in greater detail (e.g. 'where to find the necessary funding?' or 'bargaining power, time pressure, rhetoric skills' as subcategories of 'negotiations'). The implicit assumption in this typology is that a shift away from lists and labels to explications from pre- to post-measurement testifies to improved knowledge structuring. More explications on the ex post maps are interpreted as evidence that items are noted down more deliberately, less freely associative than at pre-test, leaving out items that have come to appear less relevant. We have already touched on this in the preceding sections, where we observed a reduction in listings for a number of categories from pre- test to post-test (most notably 'criteria' in the context of the burden-sharing exercise, and 'actors' and 'policy objectives' for both exercises) during the post-measurement. Since we also wanted to trace possible shifts in response types in a more systematic way, we coded every concept map according to the response type (label, explication, overly-detailed listing) that was dominant across the map. Figure 11 displays the share of concept maps for each response type at pre- and at post-test.

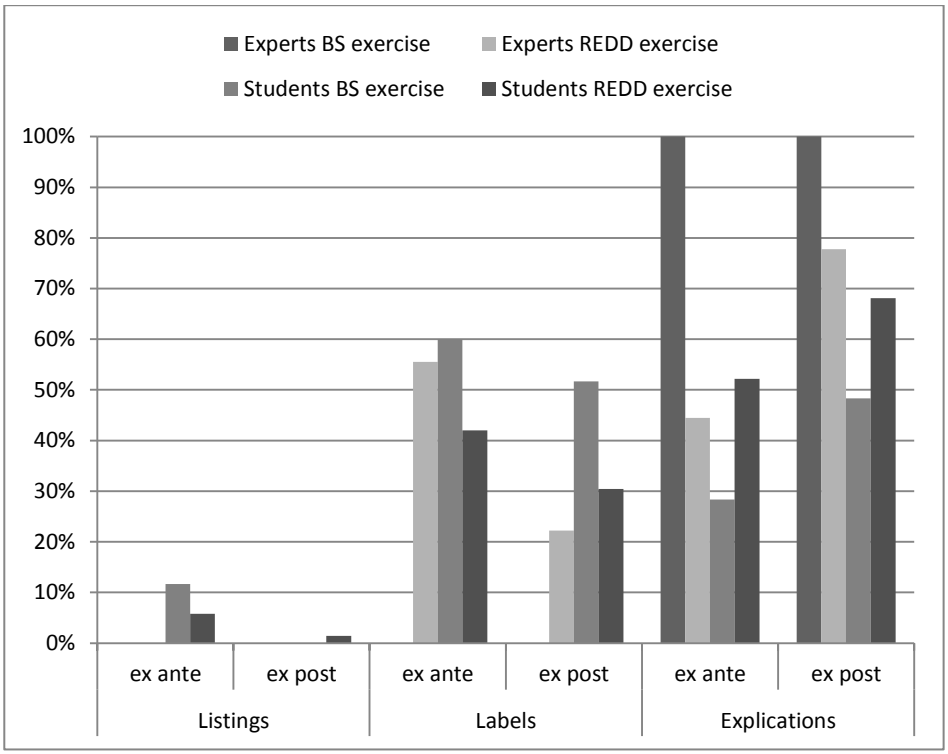

Figure 11. Share of concept maps dominated by explications, labels or lists, respectively, per participant group.

With the exception of the expert group for the burden-sharing exercise, where all available maps consisted predominantly of explications already at pre-test, for the other three groups, the share of maps in this category increased markedly. At the same time, these groups saw a decline in the number of maps that were – across all eleven concept categories – dominated by labels. The same applies for the proportion of maps containing primarily listings, which had been low already at pre-test (non-existent in the expert groups). At post-test, only a single student map (from the REDD exercise) primarily exhibited this feature. A Wilcox Signed-Rank test (treating the three response types as ordinal variables) shows that the shift from listings to labels to explications was indeed significant for the groups except the experts in the burden-sharing workshop (see Annex 2.C, Table 37).

5.2.6 Interim conclusions

In summary, analysis of the ex ante and ex post concept maps following Morine-Dershimer's method provides solid evidence that cognitive change occurred as a consequence of the policy exercise in all groups. Tracing the shifts of the concept categories from pre- to post-measurement provided valuable insights on the content and direction of this change. Across all four groups, key issues in policy design in the respective domains (interactions with the EU ETS/ design and implementation of burden-sharing schemes for the first game, and fund governance/ financing and allocation for the second) gained in salience from the pre- to the post-measurement: they were elaborated in more detail and moved closer to the centre of the map. Key themes in discussions over the course of the day thus left an imprint on participants' minds, as reflected in the design of their concept maps. This strengthened focus on 'policy content' pushed more contextual and partly more descriptive categories ('actors', 'international context', 'policy objectives' – with the exception of the latter on REDD experts' maps, where it gained in importance) more to the margin of the maps. While the renewed focus on items related to policy design on the ex post maps was a common feature across all four groups, we observed one additional phenomenon for the students taking part in the burden-sharing exercise: in this group, issues related to the politics and negotiations involved in designing burden-sharing schemes increased greatly in perceived salience, becoming the most central and second-most specific category on the ex post maps. By contrast, this category remained stable at low levels or lost in prominence in the other three groups. We will examine to what extent this trend is reflected in students' self-reported cognitive learning in the second part of this chapter.

Complementing the above analysis with metrics like the number of items and levels and especially changes in response types (lists, labels, explications) enabled a more complete assessment of the concept maps, not only in terms of shifts in the perceived salience of certain aspects, but also of evidence for improved knowledge structuring. The results point to mental maps that have become both more complex

(higher average number of items per map) and more structured (increase in the number of levels, with the exception of the experts taking part in the burden-sharing exercise, where the starting level was already very high in this respect) as a consequence of participating in the policy exercises. Comparison of the dominant type of responses from pre- to post-maps supports this conclusion. We find a reduction in the share of lists and labels in favour of more explications (again with the exception of the burden-sharing expert group, where this type dominated already at pre-test), supporting the impression of a more synthetic, evaluative outlook of the issue after the policy exercise. For both types of analysis (conceptual shifts and concept map metrics), however, the small number of available maps from the expert groups (and here especially for the EU burden-sharing policy exercise) constitutes a major limitation. Not only did it make detecting statistically significant effects harder, it also reduces the representativeness of the findings. Indeed, for the expert groups, the results from the concept map analysis are probably first and foremost proof that this method is a valid tool for tracing cognitive change also beyond an instructional or training context and that it deserves more widespread application also in other domains.

5.3 Self-reported cognitive learning

While the preceding assessment of participants' pre- and post-concept maps shed light on some aspects of cognitive learning (recorded shifts in emphasis regarding the topics of the two exercises, structuring and complexity of participants' mental maps), in the following, we aim to add a more qualitative dimension to the analysis. More specifically, in line with the expectations for cognitive learning that we outlined in the beginning of the chapter, we examine to what extent participants reported insights from the policy exercises in terms of policy substance (policy options and their feasibility in the domains of EU burden-sharing and REDD financing, respectively) and of the policy process and its dynamics. To this end, we draw on information from ex-post interviews conducted with participants about two weeks after each run as well as the post-surveys completed immediately after the exercise (or, for the students within a week after the event). The surveys included a number of closed questions on cognitive learning as well as an open-ended question that inquired what three key insights the respondent was taking home from the policy exercise.

5.3.1 Cognitive learning on policy content in the policy exercises on EU burden-sharing

Run with experts

The experts participating in the policy exercise on EU burden-sharing were overall moderately positive about the insights they gained from it in terms of policy content.

The survey responses to the question 'to which extent has the policy exercise increased your understanding of current developments in European climate policy?' reflect this (see Table 10): two respondents (14 per cent) responded 'not really', while 3 respondents (21 per cent) each chose the categories 'very much' and 'to some extent', with a further six persons (43 per cent) stating that it had enhanced their understanding 'a bit'. Given the high level of existing knowledge in the area of the policy exercise among participants, we would consider this a reasonable result, especially given the short timeframe of the exercise.

To which extent has participating in the policy exercise increased your understanding of current developments in European climate policy?

	not at all	not really	a bit	to some extent	very much
Count		2	**6**	3	3
%		14%	**43%**	21%	21%

Table 10. Self-reported learning on current policy. Experts burden-sharing exercise

The aspect that was mentioned most often in the ex post interviews was one that had received little attention in the 2008 policy negotiations on the EU Climate and Energy Package, namely the crucial importance of the criteria on which burden-sharing agreements are based. According to one interviewee, the workshop 'made me think about ways to make it (the policies and arrangements, C.H.) more flexible and what criteria you have for sharing the burden, like per capita emissions and all other ways and what was important for the countries and how you put them together - this was something new.' This theme also emerged from the open survey question: responses concerning the three key lessons learned from the event included statements like 'how important the main criteria for effort-sharing are', 'criteria for effort-sharing are paramount but divisive', 'criteria for effort sharing are very complex and yet crucial for success', and 'fairness is a social construct, interpretable in many ways.' The issue of compliance – how to ensure that member states will meet their commitments under a future burden-sharing agreement – was also raised by several interviewees, including this respondent: 'for instance compliance was new for me, as a policy-maker you have to think in a different way about the implications of your decisions.' Finally, someone emphasised that the policy exercise had prompted a more holistic understanding of the issue at hand, commenting that 'it was a very nice way to delve into the problem and see its different political dimensions.'

Participants were more sceptical, however, to what degree the policy exercise had succeeded in prompting a truly future-oriented dialogue – supposedly one of the key objectives of the policy exercise format. 8 respondents (57 per cent) responded that

the policy exercise had 'not at all' or 'not really' increased their understanding of future developments in European climate policy (see Table 11).

To which extent has participating in the policy exercise increased your understanding of future developments in European climate policy?					
	not at all	not really	a bit	to some extent	very much
Count	2	6	5	1	
%	14%	43%	36%	7%	

Table 11. Self-reported learning on future policy. Experts burden-sharing exercise.

This also came out of the interviews, with the majority expressing the view that the fact that the policy exercise was set in the future (2020) had not really had an impact on the deliberations. One interviewee maintained that moving the policy exercise into the future setting had been a good decision: 'I thought this was a good choice after all. […] It helped me a bit to reflect how thinking about climate change needs to develop further.' Another cited the 'need to balance long-term opportunity with short-term cost' as one of his/her three 'key insights' on the survey. Yet most other respondents were more negative. As somebody put it:

> *'You can get away a little bit from the everyday problems and discussions that are going on at the moment and the little things that influence the country's behaviour. But it is hard to just imagine that all these little things are not there anymore in 2020. In the end, we found out that the outcome is the same as when we had been discussing the present and not the future. To get more away from the current situation you might need something more fleshed out about what the future looks like to get people away from the present.'*

The need for more information and detail, for instance in the scenarios and role descriptions, in order to reinforce the futures dimension of the policy exercise, was also cited by other interviewees. Somebody argued that while the strategy of removing people from their current role and context by placing them into a future setting might be valuable for participants that are directly engaged in the actual negotiations as policy-makers or private sector lobbyists, it might not have been necessary for the group present at our policy exercise. He argued that, while very knowledgeable, our participants were – except for one participant –more removed from day-to-day decision-making: 'These people could easily take the role of a present-day negotiator, there was no reason to move them ten years ahead in time. […] then it is easier to stick with the present day situation because you have the facts more readily available.'

Run with students

Students rated their self-reported cognitive learning about current policy content from the policy exercise similarly, though slightly more positive (mean of 3.69 vs. 3.50), than the experts (see Table 12). 31 persons (63 per cent) stated the exercise had increased their understanding 'to some extent' or 'very much', whereas six individuals (twelve per cent) responding 'not at all' or 'not really'. The difference in score between students and experts was not statistically significant.

To which extent has participating in the policy exercise increased your understanding of current developments in European climate policy?

	not at all	not really	a bit	to some extent	very much
Count	1	5	12	**21**	10
%	2%	10%	24%	**43%**	20%

Table 12. Self-reported learning on current policy. Students burden-sharing exercise.

The insights that students reported regarding the substance of European climate policy touched on different aspects. Some coincided with issues raised in the expert group ('the notion of fairness in burden/effort sharing', 'creating and agreeing on a baseline in the context of historical emissions was very difficult'). Other statements, like the realisation that 'economic conditions in different countries need to be considered in policy formulation' or that 'national prosperity stays an important factor' probably reflect the background of the students in environmental management, with the policy exercise confronting their personal beliefs that environmental concerns should constitute a priority in political decision-making. Overall, less than one fifth of all 'key insights' that the students noted in response to the open survey question were primarily related to policy content, illustrating that this did not constitute the core of their cognitive learning during the exercise.

Contrary to the expert group, mean student scores on learning about the future of European climate policy were slightly higher than those about the present (3.88 vs. 3.69). 19 respondents (39 per cent) replied 'to some extent' or 'very much' to this question (see Table 13). Yet despite the more positive assessment regarding the knowledge gained about future developments, the fact that the policy exercise itself was set in the future seems did not seem to have resonated strongly with the students we interviewed. One commented: 'it was hard to imagine what was the same and what would be different by then. So I guess we played as if it was now.' Another was even more agnostic to this aspect: 'Now that you said it was in 2020, I wouldn't have remembered that actually. If you had asked me when it was set, I would have said this year […].' Thus, the futures context apparently had only little influence on the way the students acted throughout the policy exercise. One might

argue however that this element of the exercise was less relevant for this group anyway. As both policy-making and negotiation processes as well as the specific subject matter were rather new to most students, there was less need to stimulate 'out of the box' thinking in this manner.

To which extent has participating in the policy exercise increased your understanding of future developments in European climate policy?

	not at all	not really	a bit	to some extent	very much
Count		7	23	13	6
%		14%	**47%**	27%	12%

Table 13. Self-reported learning on future policy. Students burden-sharing exercise.

5.3.2 Cognitive learning on policy content in the policy exercises on REDD financing

Run with experts

For the policy exercise on REDD financing, we were again interested what participants reported to have learned about current and future policy on REDD, especially its financing dimension. Seven experts (58 per cent of those who completed the survey) responded that they had learned from it 'to some extent', against three persons (25 per cent) who reported that this had 'not really' been the case (see Table 14). The mean score at 3.33 was marginally lower than for the experts of the burden-sharing group (the difference was not statistically significant, however), thus to some extent disproving our expectation that the policy exercise on REDD might provide greater potential for cognitive learning.

'The REDD-ALERT policy exercise provided me with an update on current developments on REDD+ financing.'

	not at all	not really	a bit	to some extent	very much
Count		3	2	7	
%		25%	17%	**58%**	

Table 14. Self-reported learning on current policy. Experts REDD exercise.

The issues raised most often in the open survey question regarding the 'key insights' gained were – not surprisingly – connected to the core of the exercise, REDD financing. Some participants just noted 'various ways of short/medium/long term financing for REDD' or 'mechanism of financing', and one person admitted that

'raising sufficient funds will be more difficult than I thought.' Another added that 'REDD faces exactly the same financing problems as every other forestry issue. So we should look at the financing options of other international problems, especially where it has worked.' Somebody else elaborated in the ex-post interview:

> *'The discussions in policy circles and in the literature are often so devoid of flesh and blood and political context, it seems as if we just get the design [of a future REDD mechanism, C.H.] right, the rest will just fall into place and parties will live up to the commitments they made. The workshop made me realise the difficulties in raising the needed resources to an issue as uncertain and complex as REDD.'*

A similar observation was made with regard to the reality in developing countries, reflecting that what was 'missing in this discussion (and maybe the real world) is the detail on what is done on the ground and how this reflects back on things like MRV [monitoring, reporting, and verification of achieved emission reductions, C.H.], administration, funding, etc..' Other participants reported that they had understood better 'how the [REDD, C.H.] fund could be operationalised' or concluded, with regard to the planned phasing of REDD implementation, that 'readiness and actual mitigation phase should be clearly separate.' Finally, an important recurrent theme in both interviews and survey responses was increased scepticism regarding the political as well as the practical feasibility of REDD following the exercise: 'the success of REDD is not guaranteed', 'few countries actually have an incentive for an operational REDD mechanism'; 'principle is ok, but difficult to put into practice'.

The response of the experts to the second closed survey question on cognitive learning, namely to what extent the policy exercise stimulated thinking about the medium and long-term future of REDD financing, was decidedly more positive than for the burden-sharing exercise. All respondents replied that this had been 'to some extent' or 'very much' the case (see Table 15), resulting in a mean score of 4.25 out of 5. This was the highest of all self-reported learning questions in the survey across the different groups, and significantly higher than for the same survey question to the burden-sharing experts ($F(1,20.48) = 11.78$, $p < .05$) [25]. The REDD policy exercise thus seems to have achieved the objective of getting participants to thoroughly reflect about future policy developments, as illustrated by this quote of one participant: 'God knows I have been to many workshops, but this one really stands out. The long-term big picture of REDD and the related challenges and what we might do about them have become a lot more tangible.' Another added that the exercise 'did raise a

[25] The F-value is a ratio of the variability *between* groups compared to the variability *within* the groups. If this ratio is large then the p-value is small producing a statistically significant result (i.e. rejection of the null hypothesis). Df refers to the degrees of freedom. The Welch-test reported as the assumption of homogeneity of variance was violated for the ANOVA test.

lot of questions that are easily lost when one thinks only about the next steps to make REDD a reality.'

'The REDD-ALERT policy exercise made me think about the medium- and long-term future of REDD+ financing.'					
	not at all	not really	a bit	to some extent	very much
Count				9	3
%				**75%**	25%

Table 15. Self-reported learning on future policy. Experts REDD exercise.

The scenario of the REDD policy exercise moved the simulation ahead five years into the future (from 2010 to 2015). This future setting itself, however, seems again to have left only little impression with the participants. 'Placing yourself in 2015 is not so easy to do in a short time' was one reaction of an interviewee. But it did help to 'sharpen some of the conflicts we will have to expect', according to another. Overall, however, we would tend to conclude that this design element had only a very limited effect in the context of our policy exercises.

Run with students

This brings us to the cognitive learning effects from the student exercise on REDD. Unfortunately, we cannot rely on data from the closed survey questions for this group as the relevant section was omitted from the post-questionnaire due to an oversight. Nonetheless, the available qualitative data allows for relatively robust insights into the students' learning in this domain, even though it limits the extent to which direct comparisons can be drawn.

Compared to the student exercise on burden-sharing, lessons learned on policy content, which touched on various aspects related to the design and potential effectiveness and equity implications of a future REDD mechanism, figured much more centrally among the three 'key insights' that students involved in the REDD exercise wrote down in their post-surveys. About half of the items noted could be grouped under this category, as opposed to roughly twenty per cent for the students of the burden-sharing exercise (like for the other groups, the remaining items related either to the policy process and its dynamics, which we will cover in the next section, or insights into actors and attitudes, which are covered in chapter 7 on relational learning). In general, the responses covered a broad range of issues, revealing a rather sophisticated understanding of REDD and the associated political and technical challenges. Similarly to the insights reported by the experts, a large number of items dealt with aspects of REDD financing, the central theme of the policy exercise. Examples include the realisation that 'funding is difficult to find and agree upon', that

raising 'enough and stable funding for REDD+ is not a logical result of the negotiation process', that 'eligibility criteria for funds and fund management are important issues to agree upon and require stakeholder participation' or that 'funding is the most difficult part of all negotiations in general and REDD+ in particular.' A further theme, which was also reflected on the concept maps and in the outcomes on normative learning examined in the next chapter, was the importance of safeguards and co-benefits, i.e. attention to the wider social and environmental dimensions, in REDD design. Respondents stated that 'co-benefits are one of the most important outcome[s, C.H.] for developing countries for contributing to REDD+', that 'accounting for co-benefits are [is, C.H.] important to value the rights of local communities […]' and that '[s]takeholder participation is not an automatic first thought that comes to mind, but is essential for legitimacy.' A number of respondents also reflected on the institutional aspects of REDD: the exercise 'shows how much we are stuck with the current institutions like World Bank and GEF'; '[i]nstitutional aspects and governance of REDD fund will be a major challenging aspect of the scheme.' Last but not least, the open survey question prompted more general reflections on the prospects and possible impacts of REDD, with respondents reaching different conclusions in this regard. While some concluded that 'nevertheless it [REDD, C.H.] seems the best forward we have' and that 'although [despite, C.H.] the difficulties, REDD is a cheaper and more practical solution than most alternatives and therefore has the potential to be of great impact'; others warned that 'executing an agreement with such complexities and on such scale could be considered highly unrealistic' or emphasised 'the problem that if you pay the loggers not to log, how many loggers will there be.'

Compared to the other groups, the interviews that we conducted with students following the exercise on REDD financing seemed to indicate a more positive view of the decision to move the policy exercise into the future. As one interviewee commented,

> *'[i]t made a difference, because there is simply less time available to deal with the problems and there is more pressure to get to an agreement. If things are set in the future and nothing much has changed so far, then the issue becomes much larger, which is a very good thing for such exercise. And it is also good to realise that this is actually happening. Five or ten years from now we really are in that situation and it will be a shame that we haven't done more about it in the previous years.'*

Another concurred: 'I think it did [matter that the exercise was set in 2015, red]. It gave people more freedom to be creative in their thinking, since placing the situation into the future makes it an imaginary setting that is however very closely connected to reality.' Unfortunately, we are again lacking data from the relevant closed survey question to back up this observation on a wider basis and to more reliably ascertain broader cognitive learning effects about future policy for this group.

5.3.3 Broader contribution of the policy exercises - insights that are 'new to the world'?

Having concluded this discussion on individual learning about policy brings us to the broader question to what extent the two policy exercises generated insights and policy options that might not only be novel to the participants, but actually 'new to the world.' While this may be 'an excessively high standard against which to measure any study technique' (Jones, 1964, 42), it is a claim that is made in the more optimistic writings on the possible effects of simulation-games and thus warrants some examination here. Our analysis here limits itself to the expert runs, as it was mainly for these that such insights could credibly be expected.

Policy exercise on EU burden-sharing

When asked whether they reckoned that the policy exercise had generated truly innovative, novel insights or policy options, our interviewees were in the first instance somewhat sceptical. Though all but one respondent could name specific aspects of EU burden-sharing in the interview that they had not been familiar with beforehand and that thus constituted cognitive learning for them as individuals, they expressed doubts whether these really constituted 'new knowledge' at a broader level. 'No, I don't think so. I am so close to the process and follow the negotiations' was the reply of one person, supported by another, who concluded that: '[s]omebody that has been following the discussion on burden-sharing for the last ten years would not get very many new things out of it.' Having examined the minutes and outcomes of the runs and cross-checked them with our recorded observations, we would tend to agree with this assessment. Nevertheless, upon closer analysis, there were also a number of elements emerging from the negotiations and the debriefing that in our view do constitute modest advances to the general discussion on EU burden-sharing and add some sophistication to it, although it is of course very difficult to judge their innovativeness at a broader level.

The most interesting aspects in this regard probably related to the criteria or methodology for determining the emission reduction commitments of individual member states, which also featured prominently in the responses to the open survey question on 'key insights' in both the expert and the student groups. The question of criteria for burden-sharing had been hardly contested in the policy negotiations of the 2008 effort-sharing decision in the context of the EU Climate and Energy Package, which were taking place around the same time as our expert workshop. However, it turned out to be the by far most controversial topic in the policy exercise runs with both experts and students.[26] The discussions during the policy exercise illustrated

[26] The fact that the topic received limited attention in the real negotiations for the period 2012-2020 may be attributed to the Commission´s framing of its legislative proposals relating to the EU ETS, effort-sharing and renewable as a ´package´: whereas the EU ETS emphasised

that, when looking ahead to the medium-term, as was the intention with this method, the question of criteria seems set to become significantly more intractable. Since the ´low-hanging fruits´ – the relatively easily exploitable mitigation potentials in the transport and residential sectors – will have been largely used up by then in the more developed member states, a wealth-based allocation of reduction commitment appears less politically palatable than today. At the same time, mitigation in the non-trading sectors will have become costlier in the newer member states, further curbing their enthusiasm to shoulder a larger brunt of the abatement burden. One expert group sought to address this dilemma by resorting to a combined indicator that factored in marginal abatement costs, but also took account of equity considerations, both in terms of emissions and income. Assuming that member states were to accept the calculations of marginal abatement costs computed by the Commission for the whole of the Community as a basis for allocating targets , this might indeed be a feasible and innovative way forward.

A different issue that came up in the discussions was the interrelation between the EU ETS and the effort-sharing domain. As the EU ETS is expanded to include additional sectors, this inevitably reduces the mitigation options remaining in the non-trading sectors. Ensuring sufficient flexibility inside the effort-sharing domain thus becomes ever more relevant. This could be achieved either by allowing for the trading of achieved reductions among member states (with the strong performers selling on their excess emission cuts) or by relying on purchases of carbon credits from outside the Community. The latter option carries a greater risk of compromising the environmental integrity of the scheme, prompting expert group B as well as the student groups to focus on maximal internal flexibility, which, as they argued, would also increase the overall cost-efficiency of the mitigation effort. Alternatively, as was propagated by some of the new member states during the 2008/2009 negotiations, the 'firewall' separating the EU ETS from the effort-sharing domain could be softened to some extent. This would however impact the 'cap' – the centrally established emissions ceiling – under the EU ETS, with detrimental consequences for the installations covered by the trading scheme in terms of transparency and investment certainty. Both expert groups realised this risk in their discussions and as a consequence focused primarily on the first two options.

Expert deliberations in the context of our exercise further established a link that has not featured in policy discussions to date, namely the connection between burden-sharing in mitigation and in adaptation. Both Spanish teams in the two expert sub-groups argued that the relatively more severe climate impacts and related adaptation costs that their country was facing should in some way be reflected in the agreement on mitigation burden-sharing. While other teams opposed this, cautioning that such a

---

cost-effectiveness considerations, the Effort Sharing Decision squarely addressed member states' equity concerns (Haug & Jordan, 2010).

link would risk overburdening the mechanism, the policy exercise may have foreshadowed future discussions on this point. Eventually, neither group in the exercise endorsed a direct link by factoring adaptation burdens (however these might be calculated) into the determination of emission reduction commitments.

Policy exercise on REDD financing

The judgment of the expert participants that we interviewed in the aftermath of the REDD policy exercise was quite similar to the one by the participants in the burden-sharing group. Though all interviewees were positive about the insights they had gained individually about future developments and challenges related to a global REDD mechanism, they were cautious to what extent any ideas had emerged that could possibly qualify as 'new to the world'. According to one interviewee, 'it was rewarding intellectually to reflect about issues that are still two steps away from what is being discussed now in policy, but besides my broader impressions, I have a hard time coming up with a single concrete idea that emerged from it [the policy exercise, C.H.] that I would call innovative and useful.'

Our own assessment, based on minutes of the workshop and the debriefing as well as our own observations, comes to a similar conclusion. One aspect of the outcomes of the exercise that we reckoned may have pushed the boundaries of broader mainstream thinking on REDD concerned how achieved emission reductions under REDD in a country might be verified. Our expert group and also most of the student groups agreed that, following the example of the Kyoto Protocol's Clean Development Mechanism (CDM), this task should be carried out by accredited independent entities, thus according a key supervisory role for a largely public mechanism to private bodies. This choice diverged from established conceptions on this issue, which mostly consider it as a function of an independent international institution or body (Angelsen et al., 2009; Streck et al., 2009). While involving the private sector in this way may have the benefit of fully harnessing the experience with verification and certification gained under the CDM, it may also raise sovereignty concerns and risks increasing transaction costs.

Otherwise, while not producing novel ideas or policy options per se, the REDD policy exercise did provide a useful 'test drive' for two policy options related to the disbursement of payments from a global REDD fund. These had been suggested in the scientific literature and were included as options on the policy element cards in the exercise, but – as far as we can judge - had not yet reached the political discussions. The first was the suggestion to organise disbursement from a global REDD fund through an auction-based format. In such a set-up, developing countries would sell their verified REDD units to the global fund, which would first purchase the units offered at the lowest price, then those at the second-lowest, and so forth, until the provided budget is exhausted or the desired quantity of REDD units reached. In

the academic literature on the topic, Obersteiner et al. (2009) emphasise the versatility of REDD auctions, which could be designed to account for social and environmental co-benefits, or target specific groups of countries by organising auctions in separate tranches, using different qualifier criteria. However, developing country teams in both the expert and the student runs of the policy exercise fundamentally opposed the efficiency-based logic of such a system, arguing that dedicated windows under the fund catering to different purposes or priorities were both more equitable and transparent. These equity-based concerns do not bode well for the political feasibility of an auctioning approach, which is yet to enter the international climate negotiations, where the focus is currently still on the more fundamental questions surrounding REDD.

More positively received was the notion of a longer-term integration of REDD with the global carbon market. Country teams agreed that for REDD to be viable over the long run an interface of some sort with the carbon market would have to be established. Among the different options that were discussed for engineering the transition from a fund to a market-based approach, the idea of an International Investment Reserve received the greatest support in the policy exercise. An International Investment Reserve would act like an investment bank for REDD, purchasing yet-to-be created REDD units from host countries and, once verified, reselling these into the carbon markets to create returns for investors and to replenish the IIR fund (Obersteiner et al., 2009). This set-up would provide for an influx of REDD credits into the carbon markets, controlled by the reselling decisions of the Investment Reserve's governing board. The concept met with opposition from the United States team, however, who argued that the fund administration should be not be able to exert influence on the autonomous functioning of the market.

5.3.4 Insights into policy dynamics in the policy exercises on EU burden-sharing

Another aspect that we were interested in the domain of cognitive learning was the extent to which participants reported insights on the process dimension of policy-making in EU and international climate policy, respectively. Given the set-up of the policy exercises as simulated negotiations, we expected especially the students in both exercises, who we assumed to be generally less familiar with such processes, to experience cognitive learning in this domain.

Run with experts

When examining the results of the expert run of the burden-sharing exercise, however, we were surprised at the substantial level of attention to the process dimension in participants' evaluations. After all, while only a minority was directly involved in policy-making in their daily work, all had for the past years been working on climate policy issues, albeit from different perspectives. A majority of eight

participants (57 per cent) stated that the exercise had increased their understanding of the dynamics of EU climate policy-making 'to some extent' or 'very much', against two (14 per cent) and four persons (29 per cent), respectively, who opted for the categories 'not really' and 'a bit' (see Table 16). With an average score of 3.57 out of 5, this outcome was marginally higher than for the questions inquiring about other aspects of self-reported cognitive learning in this group.

'To which extent has participating in the policy exercise increased your understanding of the dynamics of EU climate policy-making?'

	not at all	not really	a bit	to some extent	very much
Count		2	4	6	2
%		14%	29%	**43%**	14%

Table 16. Self-reported learning on policy dynamics. Experts burden-sharing exercise.

The extent to which issues related to policy and negotiation dynamics had left an imprint with participants came out even more clearly in the responses to the open survey question and the ex post interviews. More than one third of experts´ reported 'key insights' fall in this group. Participants for instance highlighted the '[i]mportance of negotiation process as there will always be need to be compromise' and noted that 'it is easy to get carried away in a negotiation setting and accept text that differs from your position.' They also noted lessons that they had learned from the exercise in terms of negotiation strategies, such as the '[i]mportance of unbundling then rebundling as a way of clarifying real priorities and circumventing petty objections' or that 'the negotiating matrix needs to be very simple. There need to be a limited amount of negotiation topics that are well thought through.' Finally, reference was made to the specific policy context that the burden-sharing exercise sought to simulate: one respondent highlighted the 'limitations of EU decision-making structure' as a key insight. Yet this aspect was not necessarily novel to others; according to one interviewee, the exercise 'strengthened and reaffirmed my view of how countries operate in a negotiating setting in the EU.' Others were more sanguine, as for instance this person: 'I think it was quite a useful exercise because, especially for me with my background, it gave me some insight in the way these things work.'

Run with students

In line with our expectations, student scores on learning about policy dynamics surpassed those of the experts. 27 students (55 per cent) reported having learned 'very much' in this regard, with the average amounting to 4.35 (see Table 17). Not only was the difference in results between student and the expert group statistically significant ($F(1,61) = 8.97$, $p < .01$), the mean for learning about policy dynamics was

also much higher than for the questions inquiring about other aspects of cognitive learning in this group.

'To which extent has participating in the policy exercise increased your understanding of the dynamics of EU climate policy-making?'					
	not at all	not really	a bit	to some extent	very much
Count		1	8	13	27
%		2%	16%	27%	**55%**

Table 17. Self-reported learning on policy dynamics. Students burden-sharing exercise.

This emphasis was echoed in the responses to the open survey question, with half of all answers to the question regarding the 'key insights' from the policy exercise focusing on the process dimension. While the aspects mentioned are rather diverse, some key themes emerge, such as possible 'success factors' for negotiations: '[t]he success of negotiations depends greatly on the ability of the representative to convey their ideas in a convincing way', whereas '[t]houghtful, thorough reasoning and argumentation and underpinning is of secondary importance, considering the time pressures.' Similarly, students noted that 'negotiations in small focus groups benefits the process' and that a 'strong stance on one issue can be used as leverage against another.' Another strand of responses expressed frustration with the constraints of policy-making: 'Negotiation outcomes have no relation to objective science', 'politics is about power and dirty' and '[t]here is a discrepancy between reaching a good agreement and reaching an agreement that will actually be implemented.' Some also referred specifically to the EU context, reflecting for instance that the '[v]oting system is important obstacle to very ambitious targets.' In the ex post interviews, students also emphasised this dimension when asked what they were taking away from the exercise, as illustrated by the following quote: 'I definitely learned a lot more that day than probably in the couple of weeks of the course, really. It was fine to hear about it [how policy-making works, C.H.], but then to actually do it in practice is quite different.'

5.3.5 Insights into policy dynamics in the policy exercises on REDD financing

Run with experts

Unlike to the EU decision-making process, which provided the frame of the burden-sharing policy exercise, the exercise on REDD financing was embedded into the context of global climate negotiations. More specifically, the scenario of the exercise

stipulated that the setting of the simulation was a pre-meeting preceding the start of the actual negotiations.[27]

The REDD experts varied in their response to the question to which extent the policy exercise had enhanced their understanding of the dynamics of international decision-making, with an equal split between the four answer categories 'not really' to 'very much' (see Table 18). The mean score (3.50) was very close to that of the burden-sharing group (3.57).

'The REDD-ALERT policy exercise provided me with a better understanding of the dynamics of international decision-making on REDD+.'

	not at all	not really	a bit	to some extent	very much
Count		3	3	3	3
%		25%	25%	25%	25%

Table 18. Self-reported learning on policy dynamics. Experts REDD exercise.

Process aspects also appeared among the self-reported 'key insights' in the survey, though less dominant and elaborate than for the burden-sharing expert group. One participant highlighted the 'negotiation limitations/procedures', another simply noted to the 'negotiation process', and yet another mentioned insights into '[h]ow to phrase comments so that other delegates will react/internalise.' In the interviews, the process dimension was also brought up. According to one participant, the exercise made her

> *'[…] wonder about the value of reasoned argument versus manipulation and power in international negotiations. The exercise showed how important framing and finding the right allies is in getting an idea accepted. And how easily an objectively good option can be discarded if it is not well presented.'*

Two other interviewees, however, pointed to the fact that while the policy exercise did bear some semblance to reality in its procedures, it was unable to capture the dynamics of international climate negotiations, due to its limitation to just five countries. Based on our own observations, we would tend to concur with this assessment. Yet we maintain that, given the number of participants, the decision to play with a smaller number of teams favoured a more in-depth engagement with the

[27] The reason was the objective to maintain symmetry in design with the first policy exercise, where five parties had been represented in the simulation. For the REDD exercise, we opted for a pre-meeting to the international climate negotiations (which involve more than 180 parties) as the most credible way to justify a similar limitation in the number of participating countries.

subject matter as the teams could jointly explore the advantages, drawbacks and implications of certain options before entering into negotiations with their counterparts.

Run with students

In the open survey question as well as the follow-up interviews, the students that participated in the REDD policy exercise reported a number of key insights and learning effects concerning the process dimension of policy-making. As one student commented, 'I enjoyed it quite a lot, so I was quite excited about it. It made me more interested in that kind of topics, especially the push-pull theories of policy-making. It made also very obvious why nothing is ever binding.' Another summarised her impressions as follows: '[f]or me, it seems like the whole policy process is like a big dark hole of bargaining, agreeing […].'

Recurrent themes in the reported 'key insights' relating to policy process were the challenges involved in reaching consensus ('achieving consensus in international negotiations needs good understanding of different needs, ability to negotiate between parties and strong leadership as well'), the compromise nature of the resulting outcome ('the higher the level of consensus, the more empty words', '[c]ompromise often leads to a diluted result') and the importance of communicating clearly ('[m]iscommunication slips in easily', '[t] he importance of skilled diplomacy and effective speaking'). Yet on the whole, the policy process dimension of cognitive learning seems to have been less dominant for the students who participated in the REDD than for those taking part in the burden-sharing exercise. About a quarter of all noted 'key insights' falls in this domain (as opposed to roughly half for the burden-sharing group), and also in the interviews this aspect was mentioned rather as one among others or else discarded entirely, as for instance by this student:

> *'I didn't learn new things about negotiations and policy making since I have done things like that before. […] I learned quite a lot about REDD. I think it was interesting to see the variety of stakeholders involved in REDD. I didn't realise that so many countries have interest in REDD and would actually be willing to put money into it.'*

Substantial learning about REDD and learning about policy dynamics thus seems to have happened in more equal measure in this group.

Before drawing together the findings of this chapter, we want to probe a little deeper into the extent to which some other factors apart from the topic of the exercise and the distinction between experts and students may have influenced the extent of cognitive learning among the participants in our policy exercises.

5.4 Factors influencing cognitive learning

5.4.1 Participants' proximity to policy-making and prior expertise

The first question we were interested in was whether we could find a relation between participants' backgrounds and the extent of their self-reported cognitive learning on current and future policy as well as policy dynamics. Thus, we sought to ascertain for the expert groups whether their professional affiliation mattered for their survey scores on cognitive learning. We distinguished between four different groups: participants working in the public sector, in the private sector, in research or in non-governmental organisations (NGOs). However, since group sizes were unequal and especially since only two NGO representatives had completed the post-survey, we combined those four into just two groups: firstly public sector/NGOs, whom we assumed to be closer to day-to-day policy-making; and secondly research/private sector, whom we hypothesised might be further removed from these processes (see Annex 2.D, Table 38). Analysis of variances revealed a statistically significant difference only for cognitive learning about current policy. The fact that the participants we considered as having more distance from day-to-day policy-making reported more learning about current developments is an outcome that matches with our expectations ($F(1,24) = 6.59$, $p < .05$).

Concerning the students, we were interested if we could find differences between students of different academic backgrounds in terms of cognitive learning. The master course in Environment and Resource Management, which all students were following at the time of the exercise, draws in students from a broad range of disciplines, from engineering to anthropology and philosophy. We divided the student participants into three different groups based on their bachelor degree (natural science/humanities, social sciences and environmental studies, and finally political science/international relations), aiming to reflect the extent to which they might be familiar with the contents of the policy exercise. Unfortunately, we had again only data from the burden-sharing group at our disposal for the relevant survey questions. The descriptive statistics are reported in Annex 2.D, Table 39. ANOVA did not return any significant differences between those three categories for either aspect of cognitive learning.

In a next step, we were curious about a possible relationship between the impact of participants' existing knowledge on the topic of the exercise and their self-reported cognitive learning. The pre-questionnaires for both exercises included several questions asking participants to evaluate their expertise on a number of themes related to the content of the two policy exercises. Correlation analysis between these scores and self-reported cognitive learning effects yielded only one significant outcome each for the burden-sharing and the REDD expert group. For the former, self-reported prior knowledge on EU policy processes was negatively correlated with

learning about policy dynamics (r = -.634, p < .05). While this was in line with our expectations about learning outcomes presented in Chapter 4 – more learning on behalf of those that claimed they knew less about the topic beforehand –, the same did not apply for the results of the REDD expert group. Here, background knowledge of international climate policy – constituting the broader political context of the REDD exercise – was positively correlated with learning about current policy (r = .690, p < .05), indicating that higher baseline expertise may even have favoured cognitive learning in this regard. One possible explanation for this difference lies in the less structured nature and lesser maturity of the topic of REDD financing, where greater contextual knowledge at the outset may have helped.

5.4.2 Reasons to participate

We were also looking for potential connections between participants' reasons to sign up for the policy exercise and its normative learning effect. Evidently, this was only relevant for the expert groups since the exercise was part of the regular curriculum for the students.

In the pre-questionnaire, expert participants had been asked to evaluate the importance of a number of goals for their decision to register for the workshop in the pre-questionnaire, from primarily cognitive motivations ('I want to get an update on the latest developments in EU climate policy/ on REDD financing'), to more relational ('I want to better understand the perspectives of other actors/stakeholders') and social ones ('opportunity to network'). With one exception, there was no indication that those experts with a stronger motivation for cognitive learning also reported stronger learning effects in this domain (see Annex 2.D, Table 40). Instead, there were a number of correlations between 'softer motivations' (such as a desire to build up new relationships, interest in the research project sponsoring the policy exercise, the opportunity to network, and curiosity about the policy exercise format) on the one hand, and cognitive learning about current policy developments and policy dynamics, on the other. Hence, one might hypothesise that a more open orientation towards the aims of the day may have facilitated cognitive learning, or at least the extent to which it was self-reported by participants.

5.4.3 Impact of different roles

In a last step, we examined whether we could detect any effect of the roles that participants were playing in the exercise on the extent of their cognitive learning. For the EU burden-sharing exercise, we distinguished between three clusters of roles: wealthier EU member states (represented in the exercise by Sweden and Germany), less wealthy member states (represented by Spain and Poland), and the European Commission. Analysis of variance did not return significant differences between the three groups for either aspect of cognitive learning. We were unable to replicate the analysis for the REDD policy exercise runs, because, due to the absence of data on

self-reported cognitive learning for the student group and incomplete survey responses from the expert group, group sizes for the different roles had become too small for meaningful statistical testing.

5.5 Summary

While the first part of this chapter sought to trace recorded conceptual change by comparing pre- and post-concept maps (see Section 5.2.6 for a summary of the main findings), in the second part, we focused on cognitive learning as self-reported by participants. To this end, we combined data from survey questions (both closed and open-ended) and from post-interviews with participants. The data presented shows that some extent of cognitive learning occurred in all participant groups, but to varying degrees and with varying emphasis on different aspects. Table 19 provides a comparative summary of the main outcomes across the four groups.

Finally, in terms of measurements, the combination of different types of data – concept maps, surveys, and interviews – has proven valuable for shedding light on different aspects of cognitive learning. Analysis of the pre- and post-concept maps yielded insights on knowledge structuring as well as aggregated conceptual change from pre- to post-measurement. Interviews and surveys served to test our assumptions for some key areas where we expected cognitive learning to occur, such as policy developments in the domains covered by the two exercises, as well as policy dynamics. Recorded data from the concept maps and self-reported data from surveys and interviews thus provided largely complementary information; yet in some cases it also allowed for cross-checking insights, for instance where self-reported learning outcomes coincided with upward shifts of concept categories on the concept maps. We should however note again the limitations in terms of data availability and quality in particular for the expert groups. These have implications for the validity of the respective findings. Given the small number of observations, especially the analysis of the expert concept maps may be considered first and foremost a proof of concept, indicating that this method is suitable for generating relevant outcomes also outside of a training context, the only field where it has so far been applied.

	Policy exercise on EU burden-sharing		**Policy exercise on REDD financing**	
	Experts	*Students*	*Experts*	*Students*
Self-reported cognitive learning about policy substance	(0/+) Survey scores 'increased understanding of current policy developments' 3.50 out of 5; 'enhanced understanding of future developments' 2.36 out of 5. Insights into criteria for burden-sharing, compliance issues; Learning about future policy impeded by lack of detail; tied to present-day considerations, impact of exercise design?	(+) Survey scores 'increased understanding of current policy developments' 3.69 out of 5; 'enhanced understanding of future developments' 3.88 out of 5. Insights e.g. into importance of criteria and baselines for burden-sharing arrangements, conflict between climate and economic policy goals). Noted difficulty of projecting oneself into the future.	(+) Survey scores 'provided update on current developments' 3.33 out of 5; 'made me think about future developments' 4.25 out of 5. Insights e.g. into REDD financing options and constraints, fund operationalisation, general feasibility of REDD (by definition future-oriented issues, as REDD is only emerging/currently being negotiated). Mixed assessment of future setting ('hardly any impact' vs. 'useful').	(++ *) *no survey data available Insights into REDD financing options and constraints, importance and challenges linked to co-benefits from REDD, institutional challenges (the latter reflected by upward shifts from pre- to post-concept maps).
Insights that are 'new to the world'?	(0/+) None according to interviewees; our own analysis finds some innovative aspects in the discussions/outcomes (combined indicator for burden-sharing, implications of interrelation burden-sharing/EU ETS, attention to adaptation burden-sharing.		(0) None according to interviewees; based on our own analysis, choice of private verification bodies deviates from mainstream thinking. Test-drive for disbursement options not yet discussed in policy negotiations.	

Cognitive Learning

Self-reported cognitive learning about policy process	(+)	(++)	(0/+)	(+)
	Survey score 'increased understanding of dynamics of decision-making'3.57 out of 5.	Survey score 'increased understanding of dynamics of decision-making'4.35 out of 5.	Survey score 'better understanding of international decision-making' 3.50 out of 5.	[no survey data available]
	Qualitative data quite positive on this dimension, insights on negotiation strategies and dynamics, role of information, etc.	Constituted a key area of learning for this group (cf. reported 'key' insights and rise of politics and negotiations category on post-concept maps): e.g. negotiation strategies, smallest-common-denominator character of many consensual decisions, etc.	Some insights into nature and constraints of consensual decisions; but limited number of countries perceived as constraint to better understanding of dynamics.	Based on interview data: Insights into constraints of consensual decision-making, importance of negotiation skills, etc. Less dominant in self-reported 'key insights' than for burden-sharing student group.

Table 19. Self-reported cognitive learning, summary of main outcomes.

6. Normative learning

6.1 Introduction

This chapter examines the extent of normative learning among experts and students as a consequence of participating in the policy exercise. In Chapter 2, we defined normative learning as shifts in the views, values or paradigms that guide the decision-making processes of individuals and collectives. In our view, normative learning does not exclusively entail large scale, paradigmatic shifts, but may range from more limited alterations in viewpoint on specific issues to changes on broader, more abstract norms and value judgments.

The literature on policy games highlights the potential of policy exercises to bring about normative learning among participants (see also Section 2.3). According to Parson and Ward (1998, 133), 'simulation exercises can change the way participants think about the problem, shake up preconceptions […].' The presumed ability of policy games and policy exercises to stimulate normative learning is attributed to the communicative negotiation process at its heart, which allows participants to expose, clarify, and confront their viewpoints: '[d]uring the joint experimental action in the game, value debates become focused, sharpened and placed into operation in such a way that value tradeoffs can be negotiated' (Duke & Geurts, 2004, 312). One could say that the role-playing inherent in policy games creates a sort of two-level situation: first, participants bring to bear their personal values and opinions on the role that they are playing. The second level refers the constellation where they impersonate that role vis-à-vis the other teams. Thus, normative learning in a policy game stems is based both on having to negotiate one's own viewpoint with the role that one is assigned, and on the process of reconsidering one's own stance in light of the interactions between the different roles during the simulation.

The notion that policy games may be conducive to building consensus reflects a further line of thinking about normative learning which, while not as pervasive as in the literature on public participation, is nonetheless present also in the context of policy games. Duke (2011, 348) argues that '[o]ne of the most potent aspects of policy exercises is their ability to find common ground among competing stakeholders.' Bots and van Dalen (2007) likewise refer to 'mediation' as one potential function of policy games. They describe the opportunity to use policy games as a 'virtual negotiation table' (Bots & van Daalen, 2007, 519), which allows stakeholders to come together outside of their day-to-day environment, identify shared interests and points of convergence, and work on consensus-building. Often, role reversal – i.e. participants playing roles in the simulation that are different from their own in reality – is mentioned as a useful technique to foster such convergence (Deutsch, 1973; Lewicki et al., 1997).

Mostly, however, writings on policy games tend to recognise that consensus-building may not always be the most appropriate goal in an appraisal process, and treat the notion with caution. In general, the focus rather lies on the value of policy games in clarifying participants' values and perspectives; Bots and van Dalen (2007, 521) describe this as 'games as a parliament'. The structured negotiation process involving multiple roles helps participants uncovering previously unfamiliar perspectives of other stakeholders or better understanding the rationale behind familiar ones (which we subsume under relational learning). The assumption is that this may subsequently have a bearing on one's own normative stance, and improve the potential for cooperation.

What does this mean for our own expectations with regard to participants' normative learning in the policy exercises that we organised and ran? As we have argued in earlier chapters, given the dearth of literature on this topic, our research on the learning effects of policy exercises is necessarily exploratory. While we therefore refrain from hypotheses-testing, we did have some assumptions and expectations which we examine in more detail in the sections below. Given the focus of the two exercises, apart from more general questions on participants' normative learning in the pre- and post-survey, we mainly tested for changes in their perception of the feasibility or desirability of certain policy options and design features. We also examined to what extent normative convergence on these issues occurred among participants. The latter is perhaps not a key objective of the type of policy exercises pioneered by Toth and Parson. Yet given the emphasis on consensus building in other parts of the literature on participatory appraisal, we felt it was appropriate to investigate this dimension as well and see whether we could find evidence to that effect.

As for the measurements on cognitive learning, we also investigated the impact of the different topics of the two policy exercises as well as the differing level of participants' background knowledge on normative learning effects. Comparing the data from the different cohorts, we examined the impact of different levels of expertise on normative learning by experts and students, and also sought to draw conclusions about the relation between the different topics of the two exercises (EU burden-sharing and REDD) and their respective normative learning effect. Drawing on insights from social psychology reported in Chapter 2, which pointed to an inverted U-curve depicting the relation between existing background knowledge and propensity for normative learning, we expected to find more normative learning among the students than among the expert group. More specifically, we reckoned that the student group would, given their moderate level of background knowledge, locate itself somewhere near the top of the U-curve, whereas the experts' views and convictions might often be too engrained to be able to change significantly over the course of a single day.

Experts of the policy exercise method emphasise the value of the method specifically for complex, novel issues on which knowledge is still evolving and value judgments are still in flux (Parson, 1997; Wenzler, 1993). REDD financing has only recently made it on the international climate policy agenda and discussions surrounding it are still at an early stage. By contrast, EU burden-sharing has been on policy-makers' minds for almost two decades. While there is clearly a need for policy innovation to develop effective burden-sharing mechanisms to implement the kinds of ambitious mitigation targets which will be needed in the short- to medium term to slow down or halt climate change, one cannot speak of it as an unstructured issue. Consequently, there may be reason to expect more normative learning from the REDD exercise – at least for the experts, as for the students, both topics were equally unfamiliar.

The data we are using in this chapter mainly consists of information from the pre- and post-questionnaires that all groups completed prior to and after the policy exercise, complemented by information from the collective debriefing and ex-post interviews with participants. As in the chapter on cognitive learning, the present chapter makes a distinction between recorded and self-reported data. While the recorded, 'objective' measure is based on the shifts in response patterns on a series of propositions from pre- to post-questionnaires, the self-reported measure makes use of participants' own judgment on the extent of their normative learning in the ex post questionnaire.

This distinction also forms the basis for the structure of this chapter. We start out by describing the results of our efforts to trace normative change from pre- to post-survey based on the survey statements. We then turn to participants' self-reported normative learning, and subsequently move on to investigate the differences between the four groups, as well as the interrelation between the 'objective' (recorded) and 'subjective' (self-reported) proxy. Finally, we investigate some of the factors that may have hindered or helped normative learning among the participants of the policy exercises, based on the survey data.

6.2 Tracing normative change from pre- to post-measurement across the four exercises

As outlined above, one part of our measurements of normative learning relied on a pre- and post- comparison of participants' views on a list of items and propositions covering different aspects of the subject matter of the policy exercises, EU burden-sharing and REDD respectively. Participants completed an online pre-questionnaire one week prior to the workshop and a second one on paper, containing the same set

of items, right after the workshop.[28] Below we first describe the results of the expert and student runs on EU burden-sharing, before turning to the outcomes for the policy exercise on REDD financing.

6.2.1 Normative change in the policy exercises on EU burden-sharing

For the policy exercise on EU burden-sharing, expert and student groups alike were asked to provide their opinion on twenty items related to the politics of burden-/effort-sharing in the EU (see Annex 2.E). This part of the survey comprised three blocks: the first two blocks inquired into participants' views on the relative importance of various aspects for successfully concluding and implementing burden-sharing agreements in the EU, using a 5-point Likert scale (1 - 'irrelevant' to 5 -'very important'). Examples of factors that participants were asked to evaluate include 'compensation or 'side payments' in other policy domains', 'temporal flexibility, including options for the banking and borrowing of emission reductions' in a future burden-sharing scheme, and 'evidence of significant emission reductions globally'. The third block focused on participants' perceptions on various actors in the context of EU burden-sharing, again relying on a five-point scale (1 - 'totally disagree' to 5 - 'totally agree').[29]

Run with experts

For the expert group, we found statistically significant shifts in group means in two out of twenty survey propositions (see Annex 2.E, Table 41 and 42). At first glance, this seems quite limited because with a significance threshold of $p = 0.05$, one would expect at least one significant effect purely by chance. However, given the very small sample size, there is a considerable risk of concluding an absence of effects, whereas these may simply not be possible to measure for such small groups. Therefore, especially for the expert groups where the small-size sample problem is more acute, we focus not only on effects that are statistically significant, but also on the directions of others which fall below this threshold.

[28] For logistical reasons, this order was reversed for the student runs, where participants filled in the pre-questionnaires on paper right before the policy exercise, and completed an online survey in the week following the event. The average time lag between pre- and post-measurement remained the same between expert and student groups, however.

[29] We treated all resulting pre- and post-data as interval for the purposes of the analysis (Bryman & Cramer, 1990). From a statistics perspective, it might have been preferable to assess shifts in viewpoints by aggregating individual items. However, attempts to do so by forming a scale, either by using Cronbach's Alpha or principal component analysis, did not yield reliable results; we therefore resorted to investigating the changes for each statement separately.

The first significant shift related to aspects important for successfully implementing EU burden-sharing. On average, experts judged 'evidence of significant emission reductions globally' less relevant for the conclusion of EU-internal burden-sharing arrangements after the exercise than before. This finding, which implies that the perceived importance of international developments for EU climate policy decreased over the course of the exercise, may be explained in part by the dynamics inherent in the simulation: while the scenario sketched out a challenging future global context for the exercise, the exercise itself simulated EU-internal negotiations. The vividness of this simulation experience may have contributed to this decrease of global-level focus on the part of participants. This interpretation is supported by another comparatively large change in group means, which however does not reach the level of statistical significance: on average, participants considered 'pressure for the EU to lead internationally' a less important factor for reaching agreements on burden-sharing after the exercise than before. The other shift that reached the level of statistical significance touches upon notions of fairness and equity in the EU: agreement with the statement 'the EC protects the interests of the less developed countries too much' increased significantly from the pre- to the post-measurement. This, given how close it is to the core of the issue of burden-sharing in the EU, in our eyes represents a particularly interesting instance of normative learning.

Apart from changes in group means, we were also interested in exploring whether the pre- and post-survey statements provided evidence for normative convergence among participants over the course of the exercise. To test for such convergence, we compared the standard deviations for each item in the pre- and post-survey. Where the standard deviation decreased from pre- to post-test, we took this as a sign of normative convergence, whereas an increase of the standard deviation was interpreted as normative divergence.[30] For the expert run, we noted decreases in standard deviation (i.e. normative convergence) from pre- to post-measurement for only eight out of the twenty survey statements, while it increased in ten cases, and remained stable in the remaining two (see Annex 2.E, Table 42). These findings do not lend support to the hypothesis that our policy exercise induced normative convergence among the group of participants. However, it is noticeable that the order of magnitude for the increases in standard deviation (indicating normative divergence) is very small in all cases (with a maximal value of $\Delta SD = .18$ for a scale ranging from 1 to 5) and may thus not be particularly meaningful. By contrast, the largest instance of normative convergence amounts to $\Delta SD = .47$ (for a total scale of 5), respectively, providing a slightly firmer indication that group opinion converged on the need for a 'strong role of the Commission as an impartial facilitator.' The

[30] Since a normal distribution cannot be assumed for the data set of pre- and post-standard deviations and since there is no non-parametric test for paired samples, our analysis relies on assessing the visible changes in standard deviations between the pre- and the post-questionnaire.

following section analyses normative change for the student runs of the burden-sharing policy exercise, based on the same set of pre- and post-survey statements.

Student runs

As with the expert group, we were interested in detecting shifts in group means as well as indications of normative convergence or divergence among the students taking part in the policy exercise on burden-sharing.

Overall, the picture emanating from the student questionnaires resembles the findings from the expert run (see Annex 2.E, Table 43 and 44). At a $p < .05$ level of statistical significance, we find significant shifts in group means for two out of twenty statements, with two further shifts significant at $p < 0.1$. Interestingly, these changes concern different statements than in the case of the expert exercise. For instance, students rated the importance of 'arrangements which take account of historic emission reductions' significantly higher after the exercise than before. This reflected the controversy surrounding the setting of the base year (the year from which emission reductions are calculated) in the simulation run, and also resonated with some of students self-reported 'key insights' in the post-survey (see Chapter 5). The expert group had probably been more aware of the difficulties this issue had provoked at earlier policy negotiations and had thus rated it higher on average already at pre-test (M_1 = 3.19). The second significant shift concerns the role of the poorer EU member states in concluding burden-sharing agreements, whose 'willingness […] to make climate change a political priority, even if this comes at a certain economic cost' students considered on average more important after the exercise than before. The two statements that showed shifts in means significant at the $p < 0.1$ level related to the importance of a 'strong role of the Commission as an impartial facilitator' and to the need for 'temporal flexibility, including options for the banking and borrowing of emission reductions'. The former is one of the statements where we found relatively strong normative convergence in the expert group, indicating that at least some of the dynamics in pre- and post-surveys revolved around the same themes in both groups.

Turning to the question of normative convergence in the student exercise, we note that on the whole, changes in standard deviations are quite small for this group, amounting to a maximum score of ΔSD = .20 (by comparison, the largest change in the expert group was ΔSD = .47). There are as many (slight) instances of normative convergence as there are of normative divergence. Overall, the data does not provide any firm evidence whether the group's viewpoints moved closer together as a consequence of the exercise or whether the opposite was the case.

6.2.2 Normative change in the policy exercises on REDD

Expert run

Replicating the design for the burden-sharing questionnaire, pre- and post-questionnaires prompted participants in the REDD policy exercise to evaluate a number of statements relating to the potential and prospects for a global REDD fund on a 5-point likert scale (1-totally disagree to 5-totally agree). The eight statements aimed to gauge participants' viewpoints on a global fund as an interim financing vehicle for REDD, its ability to generate sufficient funding for REDD, the question of geographical balance in REDD and the issue of environmental and social co-benefits in REDD accounting, among others. The survey also contained an additional section which asked participants to rank seven items based on the level of controversy they thought these might generate in future negotiations on a global fund for REDD. Each rank could only be assigned once, thus forcing participants to establish an ordinal ranking between the items.

Probably also related to the small sample size (n=8) for the expert run (see the discussion on this above), t-tests per statement/item did not reveal any significant shifts in group means from pre- to post-test neither for the first part (see Annex 2.E, Table 45 and 46), nor for the second part of the questionnaire (see Annex 2.E, Table 47 and 48). The only case approaching statistical significance at $p < 0.1$ was a statement prompting a more general evaluation of the prospects for REDD ('establishing a REDD+ mechanism that will effectively reduce tropical deforestation is almost impossible'). On average, agreement with this proposition, while still low after the exercise, was clearly higher than before. This is an interesting finding since it provides some – albeit weak – evidence that the policy exercise led some participants to re-evaluate their views on the chances of success for REDD – a clear, though limited, instance of normative learning.

Looking more closely at other – not statistically significant – shifts in means for the ranking exercise, we note that the issue of fund capitalisation and replenishment, as well as disbursement criteria, monitoring requirements and accounting for the co-benefits of REDD were on average judged more controversial ex post than ex ante. Deciding on the modalities for verifying developing countries' REDD performance, by contrast, was on average considered less controversial after the exercise than before. This probably resonates with participants' experience during the policy exercise: the difficult negotiations on REDD funding during the simulation underscored the intractability of this issue, whereas verification of performance was one of the few areas where participants were able to reach a relatively concrete outcome at the end of the day (see Chapter 4).

Turning to the question of normative convergence, we again do not find much evidence for consensus-building in the group as a consequence of the exercise. In fact,

the trends in standard deviations differ between the two relevant parts of the questionnaire. Participants' scores did converge to some extent in the ranking task, with standard deviations decreasing for five out of seven items. However, for the second part of the questionnaire – eight propositions on prospects and design for a global REDD fund – we find increases in standard deviations (indicating normative divergence) for six out of eight cases. Looking at the largest shifts within this set, we note that while participants' views seemed to converge to some extent on the value of a REDD fund as an interim step before linking REDD to the global carbon market, the spread of opinions increased considerably on the question to what extent co-benefits should be made a key criterion for disbursing funds. Coincidentally, discussions on co-benefits in REDD were quite heated during the policy exercise and may have made some participants more aware of the trade-offs involved in enforcing higher environmental and/or social standards for REDD activities.

How to interpret these outcomes? The occurrence of some degree of normative convergence for the ranking part of the survey can be explained through the common experience of day-long negotiations, which impacted impact participants' views of the perceived saliency of the various issues. There is also a methodological conclusion to be drawn from this. Such a ranking exercise may be useful for gauging to what extent views on the importance of the various issues converged. However, tracing pre- to post-shifts in rating the propositions, as we did in another part of the survey, may in hindsight be the stronger indicator for genuine normative learning. Therefore we would tend to attribute more weight to the results regarding the propositions which seek to gauge participants' more fundamental convictions about the goals and perceived effects of REDD as a policy instrument. Results for this section point to diverging rather than converging viewpoints among the expert group and thus run counter the often-held assumption that policy games and policy exercises are generally conducive to consensus-building. This was also our impression at the debriefing, where opinions continued to be far apart on a number of fundamental points, including the ultimate objectives of REDD (should it focus on carbon alone or seek to embrace the problem of deforestation in all its complexity?) and realistic sources for financing.

Student run

The students participating in the REDD policy exercise completed the same questionnaires as the experts. Unfortunately, confusion among some of the students regarding the ranking exercise in the paper-based pre-questionnaire made some recoding for this part of the survey necessary.[31] Looking at the pre- and post-results

[31] Rather than assigning each rank only once among the items listed, about a third of the respondents assigned several ranks multiple times, e.g. by according the first rank ('most

of the ranking, we find that, based on the recoded data (see Annex 2.E, Table 49 and 50), there is a significant change group in means from pre- to post-test for two out of seven items, namely 'disbursement criteria' and 'institutional arrangements'.[32] The heightened attention for the importance of institutional arrangements after the policy exercise corresponds with our observations during the runs. Over the course of the exercise students became increasingly aware of the deeply political character of any decision on institutional issues as well as its implications, leading to drawn-out negotiations on this issue in most sub-groups.

T-tests for the second part of the survey, the propositions on REDD, returned just one significant result (see Annex 2.E,

Table **51** and 52). Students' assessment of the statement that co-benefits should become a key criterion for disbursing REDD funds, were on average significantly less positive after the policy exercise than before. Again, this resonates with our observations: going through the exercise, students developed a better understanding of the difficulties that stringent standards for co-benefits of REDD might present in particular to the participation of least developed countries in a REDD mechanism. This may have caused them to view this option more sceptically than at the outset. Interestingly, the co-benefits aspect is also one where cognitive and normative learning effects intersected for this group as this theme gained in salience on participants' concept maps and was also frequently cited as a self-reported key insight (see Sections 5.2.4 and 5.3.2). A second shift significant at a $p < 0.1$ level occurred for a statement cautioning of a potential 'institutional lock-in' that a global REDD fund might entail and that might delay or prevent the transition to integrating REDD into the global carbon market. On average, students found this proposition more pertinent after the exercise than before. This may have to do with greater realism about potential funding constraints and other shortcomings of a fund-based instrument following the exercise.

Finally, to what extent do we find evidence that the exercise gave rise to normative convergence, or, on the contrary, normative divergence among the students? On the whole, evidence for alignment of viewpoints in this group was even weaker than in the expert run. Contrary to the expert group, where a degree of normative convergence had been noticeable for this part of the survey, for the student group, we note four instances of normative convergence and three of divergence across the

controversial') to two policy issues at once. In order to avoid losing the information in the 'faulty' responses, we recoded them, based on the logic laid out in Annex 2.F.

[32] The effect for the item 'institutional arrangements' was also significant in the original data set, excluding the 'faulty' response sets, thus lending additional support to its robustness.

set of seven items (see Annex 2.E, Table 52). In addition, the changes are for the most part quite small. As for the propositions on REDD, there is not a single one where the standard deviation decreased from the pre- to the post measurement. The changes are again minor in nature. However, they do lead us to tentatively conclude that – as for the expert run – the REDD policy exercise seems to have prompted students' individual viewpoints on REDD to diverge rather than converge.

6.2.3 Summary

Having thus described the outcomes on recorded normative learning from all four policy exercise runs, what do we make of the results? On the whole, the patterns are quite consistent across the four runs. Dependent-sample t-tests picked up only a few shifts in group means for the statements included in pre- and post-questionnaires. We did find slightly more significant shifts in means for the student exercises, which might be interpreted as an indication for more – though by no means strong – normative learning on their part. Regarding the question of normative convergence within the groups over the course of the exercise, there is hardly any evidence that this did indeed occur. On the contrary, for the REDD simulation, we even have some indications that views diverged more afterwards than beforehand.

Turning to the differences between experts and students discussed above, statistically significant shifts in means and noticeable changes in standard deviations sometimes concerned the same items for both groups, sometimes different ones. This is not entirely surprising: after all, both groups underwent the same intervention, shared similar experiences and drew to some extent similar conclusions from it. Yet at the same time, given the differing starting points in terms of background knowledge between expert and student groups, some 'eye-openers' for students, which gave rise to normative changes in this group, were probably well-established facts for the experts already before the exercise. This impression was also confirmed during ex-post interviews with both groups.

In sum, the recorded data provides no evidence for strong normative learning through the policy exercise. One should keep in mind, however, that, especially for the expert groups, we may not have been able to detect medium-sized or small effects due to the limited sample sizes. Another explanation may be the limitations that such tests present for detecting normative change in the first place. There is a risk that respondents may not genuinely reconsider the content of the propositions in the ex post questionnaires, especially if the time interval between pre- and post-test is short. Deyle and Schively Slotterback (2009) emphasise an additional difficulty in this regard, based on their own experience with such measurements: in their view, especially for open forms of interventions like participatory processes, it is very difficult to precisely pinpoint in advance the areas where normative change might occur, and to formulate the propositions accordingly. This may be easier in more

structured contexts, for instance where the goal is to evaluate the impact of a training session providing a specific set of information to respondents. And yet in our view this does not imply that researchers should not pursue such ventures, but rather that there is a strong case for triangulating quantitative and qualitative, recorded and self-reported data in order to increase the robustness of the results (see also Chapter 8). In the following, we compare findings from the recorded data discussed above with those self-reported by participants based on an explicit survey question to this effect.

6.3 Self-reported normative learning and its relation with recorded normative change

6.3.1 Self-reported learning and between-group differences

Post-questionnaires for both policy exercises included a prompt asking participants to evaluate to what extent the policy exercise had changed their viewpoints on EU burden-sharing and REDD financing, respectively. Table 20 summarises the outcomes for all four groups.

To what extent did the policy exercise change your viewpoint on EU burden-sharing/REDD financing?

			not at all	not really	a bit	to some extent	very much	margin total
Experts	EU Burden-sharing	Count		6	4	4	1	15
		%		**40%**	27%	27%	7%	100%
	REDD	Count		1	7	3	1	12
		%		8%	**58%**	25%	8%	100%
Students	EU Burden-sharing	Count		9	16	19	5	49
		%		18%	33%	**39%**	10%	100%
	REDD	Count	1	6	23	34	13	77
		%	1%	8%	30%	**44%**	17%	100%

Table 20. Survey responses to post-questionnaire question on normative learning.

We find that, among the four groups, participants in the expert group on EU burden-sharing generally reported the lowest levels of normative learning. Six individuals, 40 per cent of the respondents, indicated that the policy exercise had 'not really' changed their viewpoint on the topic at hand, whereas the rest said it had changed it 'a bit' (4 respondents or 27 per cent), 'to some extent' (4 respondents or 27 per cent)

or 'very much' (1 respondent). The scores were slightly higher for the expert group taking part in the REDD exercise, with only one respondent indicating he had 'not really' learned in normative terms, while all others were more positive in their evaluation. As expected, the picture is more positive for the student groups. 44 per cent and 17 per cent of the students taking part in the REDD exercise indicated that their views had changed 'to some extent' and 'very much' as a consequence of their participation. For the burden-sharing exercise, the student survey results are only slightly lower, with 39 per cent and ten per cent choosing the two highest categories.

We sought to explore statistically the variations in self-reported learning between the four participant groups. More specifically, we were interested as to whether participants' backgrounds or the topic of the exercise made a difference for the reported learning effects. To this end, we used analysis of variances, with group affiliation (experts vs. students) and topic of the exercise (EU burden-sharing vs. REDD) as independent variables.[33] Figure 12 below shows the unweighted means of the four groups for this model (for a scale of 1 – 'not at all' to 5 – 'very much').[34]

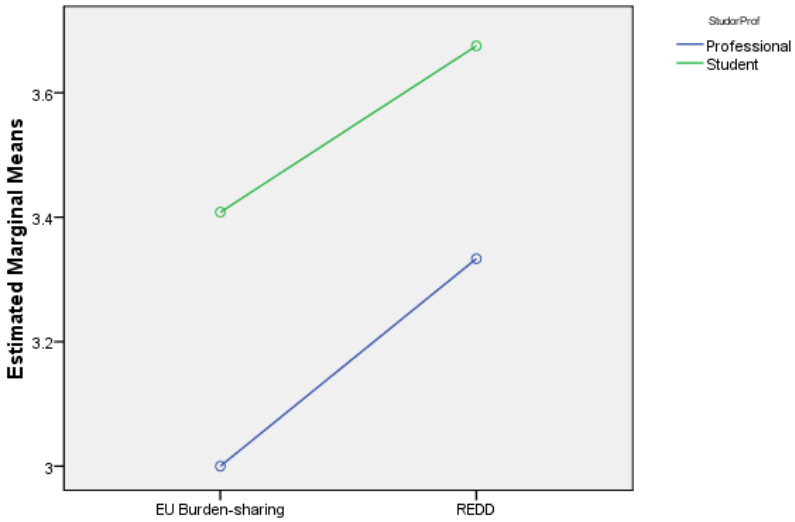

Figure 12. Self-reported normative learning on ex-post questionnaires, estimated marginal means for the four participant groups (based on a 1 – 'not at all' to 5 – 'very much' interval scale). The

[33] Within the REDD student group, we excluded one outlier, who – as the only individual in our dataset – had reported that he had 'not at all' changed his viewpoints over the course of the exercise (yet whose responses to the survey statements tracing normative change had shifted more than the group average).

[34] Unweighted means (also referred to as 'estimated marginal means') are appropriate here because the four samples sizes are unequal. In such cases, using unweighted means eliminates the possibility of confounding the effects of different variables.

four data points represent the estimated marginal means for the four groups of participants – the groups on EU burden-sharing on the left, the groups on REDD on the right. The lower blue dots represent the expert groups, and the green dots higher up the student groups. The lines serve to compare between-group means at pre- and post-test.

As Figure 12 indicates, these mean scores are higher for the REDD exercise than for the burden-sharing one, both in the expert and the student group. At the same time, there is a notable difference between the levels of normative learning by students and by experts – they are much higher for the former than for the latter. Looking at the outcomes of the ANOVA, we see that the model as a whole is significant ($F(3) = 2.81$, $p < .05$), which means that the two factors (group affiliation and topic of the exercise) help explain the difference in the reported learning effects between the groups. One of the two main effects, group affiliation, is significant at the $p < .05$ level. There was thus a significant difference between self-reported normative learning among the two groups, experts and among students ($F(1,148) = 4.36$, $p < .05$). The second main effect, the topic of the exercise, reached statistical significance at the $p < .10$ level ($F(1,148) = 2.90$).

Testing separately for the differences in normative learning between experts and students on the one hand, and EU burden-sharing and REDD on the other, both effects proved significant. The topic of the exercise thus had a significant impact on its learning effect ($F(1,151) = 4.53$, $p < .05$), as had participants' group affiliation ($F(1,151) = 4.84$, $p < .05$). These differences correspond to the expectations laid out in the introduction to this chapter. We have indeed found more self-reported normative learning among students, whose more limited background knowledge may have made it easier for them to revise their views over the course of a daylong intervention. And the relatively novel issue of REDD financing seems to have sparked more normative learning in both groups than the topic of EU burden-sharing.

6.3.2 The interrelation between the 'subjective' and the 'objective' measure of learning

Having examined the results for both recorded and self-reported normative learning, we were interested in the degree of congruence between the two measures. Given the call for using different types and source of data to enhance the robustness of an evaluation (see Chapter 3), this comparison seemed a natural extension of the approach. In order to compare them, we first had to develop an aggregated measure of recorded change based on the pre- and post-survey statements, which we had only analysed individually in the previous sections. In order to combine all these items into a single measure, we calculated the average change per participant across all survey statements from pre- to post-measurement, thus devising an overall score

of recorded normative change for each individual.[35] Table 21 below shows the means and standard deviations for these aggregated measures.

		Mean shift in score per participant from pre- to post-survey	**Standard deviation**
Experts	EU Burden-sharing	.55	.17
	REDD	.81	.40
Students	EU Burden-sharing	.65	.23
	REDD	.82[36]	.28

Table 21. Recorded normative change, means and standard deviations.

We see that, as with self-reported normative learning, average shifts per participant from pre- to post-questionnaire are bigger for the policy exercise on REDD than the one on burden-sharing. They are also greater, although only marginally so, in the case of the REDD exercise, for students than for experts. As with the self-reported learning data, we used factorial two-way ANOVA (with group affiliation and topic of the exercise as independent variables) to explore these differences in greater detail. Figure 13 below depicts the estimated marginal means for recorded normative change in this model.

Again, the model as a whole proved statistically significant ($F(1,132) = 10.68$, $p < .05$), indicating that the two factors (participants' background and the topic of the exercise) help explain the variation in mean scores for recorded normative learning. Yet the homogeneity of variances, an important assumption when conducting ANOVAs, was violated for this test, and attempts at data transformation did not remedy this.[37] Given the fact that ANOVA is generally quite resilient to violations of this assumption (Hill & Lewicki, 2006), however, this finding may still be meaningful. Since the topic of the exercise seemed to be driving most of the variance in our sample, we ran an additional t-test, which provides an option that does not assume equal variances

[35] We used only the propositions on REDD, and not the REDD ranking exercise, to calculate these shifts for the REDD exercise. This was due firstly to the fact that the scale for the REDD rankings spanned 1-7, whereas it was 1-5 for the other parts of the survey. While we could have adjusted for this, we felt that this part of the survey was also somewhat different in nature and thus perhaps not directly comparable to the other sections.

[36] The mean shift per item for the outlier, who reported that he had 'not at all' changed his viewpoint, was 1.07, above average even for this group, which had the highest values for the 'recorded change' proxy.

[37] Homogeneity of variances refers to the expectation that the variances of the groups that are measured are equal. If the variances are unequal, this can affect the Type I error rate, i.e. lead to an incorrect rejection of a true null hypothesis, thus yielding a false positive.

between groups, to further investigate the difference in means between both exercises. The t-test showed that the recorded normative shift from pre- to post-questionnaire was significantly different at $p < .0001$ for the policy exercise on EU burden-sharing ($M = 0.63$, $SE = 0.03$) and the one on REDD financing ($M = 0.81$, $SE = 0.03$, $t(128) = -4.30$, $p < .0001$), thus supporting the finding that participants in the REDD exercises learned more than those participating in the burden-sharing events.

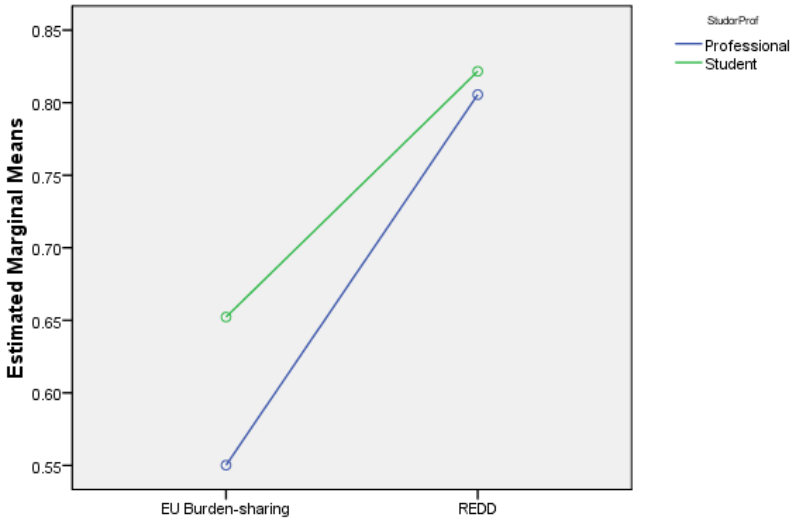

Figure 13. Recorded normative learning (mean shift per survey item from pre- to post-test), estimated marginal means for the four participant groups. The lower blue dots represent the expert groups, and the green dots higher up the student groups. The lines serve to compare between-group means at pre- and post-test.

Coming back to the interrelation between the self-reported and the recorded measure of normative change across the four policy groups, Figure 14 provides some initial insights. We see that on average, the proxy for recorded change remained fairly stable no matter whether an individual responded that his or her viewpoint had changed 'not really', 'a bit' or 'to some extent' as a consequence of the exercise (the data point on the left represents the outlier from the REDD student exercise). Only those who reported that they had learned 'very much' in normative terms also had significantly higher proxies of recorded change (i.e. their responses to the statements in pre- and post-questionnaires had changed more than the average). This is reflected in the correlation coefficient for the two measures: they are positively, although by no means strongly, correlated ($r = .209$, p (one-tailed) $< .01$) for the complete set of participants. There is thus some kind of relationship between the two measures, yet it is not linear.

Before we turn to drawing together the findings of this chapter, we want to probe a little deeper into the extent to which some other factors apart from the topic of the exercise and the distinction between experts and students may have influenced the extent of normative learning among the participants in our policy exercises.

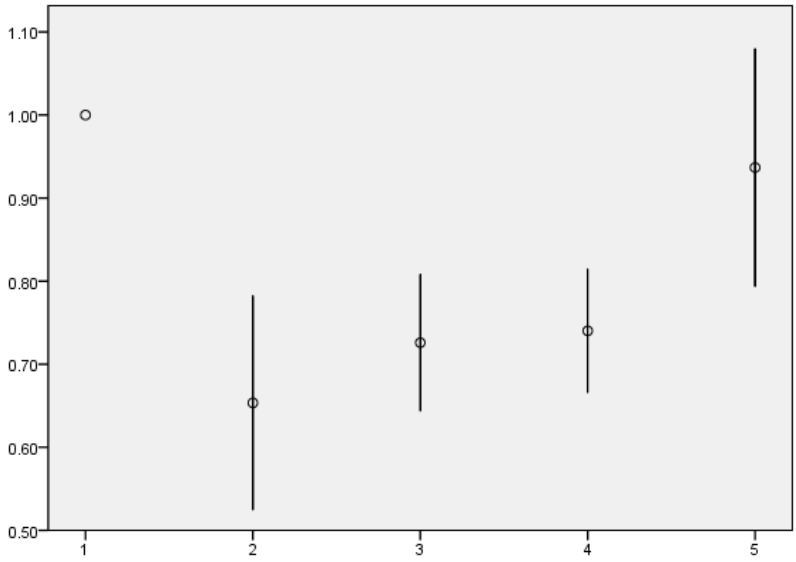

Figure 14. Interrelation between recorded and self-reported normative change. The x-axis refers to the Likert scale used in the self-reported question as to how much each participant changed their views on the topic of the exercise (1 = 'not at all' to 5 = 'very much'). The y-axis shows the mean and standard deviation of the proxy for recorded normative change. The latter is calculated as the mean change across all statements per participant.

6.4 Factors influencing normative learning

In our examination of possible factors that may have influenced participants' normative learning, we followed the approach used in the chapter on cognitive learning (see Section 5.4). We first investigated the possible impact of participants' background and expertise, before moving on to their reasons for participation and the roles they had assumed during the policy exercise.

6.4.1 Participants' backgrounds and prior expertise

In a first step, for the expert groups, we were interested whether participants' professional affiliation might help explain their scores on self-reported or recorded normative learning. However, statistical tests did not pick up any differences between the four professional groups (public sector, the private sector, research or NGOs), neither in terms of their self-reported normative learning nor for recorded normative change. The same applied when distinguishing between only two groups

who we assumed might be closer or more removed from day-to-day policy-making in their daily occupation, respectively (see Section 5.4.1). Professional affiliation thus does not seem to have had a strong impact on experts' normative learning. It is possible however, that the small sample sizes and the unequal size of the groups (with a larger share of participants coming from research and academia) prevented us from detecting smaller effects.

For the student groups, we were interested if the different academic backgrounds of the participants in the Master's course in Environment and Resource Management mattered for self-reported and recorded normative learning effects. Like for cognitive learning, we distinguished between Bachelor's degrees in natural science/humanities/others, social sciences and environmental studies, and finally political science/international relations, seeking to reflect the extent to which they might be familiar with the contents of the policy exercise. Interestingly, analysis of variance did not reveal any significant differences between those three categories, neither for self-reported learning nor for the normative change proxy.

If these general differences in participants' backgrounds did not significantly influence normative learning, how about the impact of participants' existing knowledge on the topic of the exercise? The basis for the latter was their self-assessment in the pre-questionnaire, where they had been asked to evaluate their expertise on a number of themes more or less closely related to the content of the two policy exercises. Not surprisingly, we found that for the first policy exercise, self-reported normative learning was negatively correlated with (self-reported) background knowledge on burden-sharing issues, the focus of the simulation (r = -.332, p < .05). Similarly, there was a negative correlation between expertise on climate and energy law and the proxy for recorded normative change for the first policy exercise (r = -.316, p < .05). This again lends support to the assumption of an inverted U-curve depicting the relationship between pre-existing expertise and normative change. According to the U-curve proposition, while an intermediate level of expertise facilitates normative learning, possessing expert knowledge in a domain actually reduces the chances for normative change. The data from the REDD exercises supports this: self-reported normative learning was negatively correlated with self-reported expertise on international climate policy (r = -.255, p < .05) and on development finance (r = -.301, p < .01), respectively.

6.4.2 Reasons to participate

In a next step, we were looking for possible connections between participants' reasons to sign up for the policy exercise and its normative learning effect. In the pre-questionnaire, expert participants had been asked to evaluate the importance of a number of goals for their decision to register for the workshop in the pre-questionnaire, from primarily cognitive motivations ('I want to get an update on the

latest developments on REDD financing'), to more relational ('I want to better understand the perspectives of other actors/stakeholders') and social ones ('opportunity to network').

We found for both policy exercises that there were rather strong correlations between the latter, 'softer' motivations to participate in the exercise and self-reported normative learning. Thus, for the burden-sharing group, self-reported normative learning was positively correlated with interest in the project (r = .609, p < .05), opportunity to network (r = .624, p < .05), and the format of the workshop (r = .625, p < .05). For the REDD exercise, the same applied for interest in the project (r = .740, p < .05), and the desire to 'get a better grasp on who is involved in this issue and to build up relationships' (r = .674, p < .05). The only significant correlation between cognitive motivations to join the policy exercise and normative learning concerned the aggregated measure for recorded change. For the burden-sharing group, this was significantly correlated with the desire to learn about current developments in European climate policy (r = .751, p < .05) and the wish to better understand the dynamics of EU climate policy-making (r = .726, p < .05).

We can conclude from the above that those who attended the workshop for primarily social or relational purposes were also more likely to report having learned from it in normative terms. Recorded change on the other hand was connected – at least for the burden-sharing group – to cognitive motivations. Obviously, these findings are very tentative given the small sample sizes and exploratory nature of this research, but may nonetheless provide material for hypothesis-building in future work, an issue which we will discuss in more depth in Chapter 8.

6.4.3 Roles in the policy exercise

Finally, could we find any connection between the roles that participants assumed during the policy exercise and the degree of their normative learning? Again, we replicated the approach we used in Chapter 5 on cognitive learning, differentiating between three clusters of roles for the burden-sharing policy exercise: wealthier EU member states (represented in the exercise by Sweden and Germany), less wealthy member states (represented by Spain and Poland), and the European Commission. Analysis of variance did not return significant differences between the three groups, however, neither for self-reported nor for recorded normative learning.

In the case of the REDD exercise, we made a distinction between those representing developing and those representing developed countries. In this case, recorded normative change was significantly higher for the participants that played developing rather than developed countries during the exercise (F(79,1) = 5.50, p < .05). It thus seems that, given the predominance of Northern participants in our exercise, those

who represented a country further removed from their own realm of experience also learned more in normative terms.

6.5 Summary

This chapter sought to ascertain the extent of participants' normative learning over the course of the policy exercise runs. Given the subject matter of the exercises, we focused primarily on the feasibility or desirability of certain policy options and policy design features in the context of EU burden-sharing and REDD, respectively. As for the analysis of cognitive learning effects (see Chapter 5), we relied on both recorded and self-reported measurements. Pre- and post-questionnaires constituted the primary source of data for this purpose.

We first sought to trace normative change by measuring to what extent participants' assessment of a series of closed survey items shifted from pre- to post-test. On the whole, we found only few statistically significant changes in group means across all four groups of participants. The items on which group means shifted significantly varied from group to group, indicating that normative change, where it occurred, does not necessarily seem to have revolved around the same issues between expert and student groups. Since some writings on policy games point to consensus-building as one potential function of policy games, we also looked for indications of normative convergence within groups, measured as reductions in the standard deviation from pre- to post-test. We found hardly any evidence to this effect, but should also note that it is not a key goal of the type of game we organised.

The picture emanating from participants' self-evaluation of their normative learning, based on a survey prompt to that effect, was slightly more positive. Analysing the differences between the different groups of participants, our expectations in this regard were for the most part confirmed: based on the self-evaluations, REDD as the more unstructured, novel topic seems to have given rise to more normative learning than the exercise on EU burden-sharing. Likewise, students reported higher normative learning effects than experts, in line with the 'inverted U-curve' theory which assumes that a moderate level of background knowledge facilitates the revision of pre-existing viewpoints. The difference between the 'REDD group' and the 'EU burden-sharing group' was significant also for the recorded measurements for which we had constructed a measure of aggregate normative shifts. Determining into the impact of other factors (participants' background, existing knowledge, roles) turned out more to be more difficult.

7. Relational Learning

7.1 Introduction

Better understanding of others' positions, mind sets, and preferences, greater trust and an enhanced ability to cooperate make up the key elements of the notion of relational learning as defined in Chapter 2. While all these aspects can be relevant in different types of appraisals, a review of the literature on simulation-gaming for policy development yields a more limited set of relational learning effects potentially relevant for this group of approaches. After all, policy games are mostly shorter interventions, which last for only a day or two. By contrast, other types of policy appraisal such as public participation processes often involve a series of meetings over a longer timeframe, and thus provide greater opportunity to build new or improve existing relationships. Likewise, many appraisals draw on an established, cohesive network of participants. In many policy games, by contrast, including the expert runs of our own policy exercises, participants do not know each other beforehand, come from different places (in our cases, also different countries), and, while working on the same issues, will not necessarily cross paths again in the near future. This contributes to diversity in terms of the backgrounds and perspectives represented in the group, which is often cited as a key condition for cognitive and normative learning. Yet at the same time, it inevitably limits the possibilities for detecting changes in the relationships within the group.

After this initial disclaimer, what are the relational learning benefits evoked in the literature on policy games and policy exercises? The contribution that these methods can make towards improving participants' understanding of others' mind sets is the aspect that is mentioned most often in this regard. Authors agree that policy games can help clarify and sensitise participants to others' values and world views (Underwood & Duke, 1987; Van der Meer & Geurts, 1995), also due the fact that 'when playing a role, positions and opinions can be magnified and identified more easily' (Bots & van Daalen, 2007, 521). By having to argue in favour of or against certain policy options and seeking to convince opponents of one's own point of view, 'the underlying objectives of parties crystalize, surface and become more explicit … The potential of a game for assisting individuals in viewing the situation from other perspectives through playing other roles is a major benefit of this approach' (Underwood & Duke, 1987). This last point made by Underwood and Duke draws attention to the dual nature of relational learning in policy games: they provide participants with the opportunity to learn not only about each other's mind sets, but also about the values and interests underlying the roles that they or others are playing during the policy game or exercise. Naturally, the gap between these two dimensions increases or decreases depending on how close the roles are to participants' own background and position in real life (a dynamic which also applies

for normative learning, see Chapter 6). Yet the fact that 'games aid the strategic understanding of the actor-network configuration' (Mayer, 2009, 849) remains an undisputed benefit of this method. Some authors (Duke & Geurts, 2004; Klabbers et al., 1995) also report improved communication in the group of participants as an outcome of running a policy game, 'which, over the longer term, could lead to a more solid foundation for action' (Klabbers et al., 1995, 131). However, they neither go into further detail nor present supporting evidence for this claim.

Looking specifically at policy exercises, it seems that, given the strong cognitive motivations underlying this method, which emphasise putting the right conditions in place for enabling learning about the policy problem and developing and appraising policy options, relational learning objectives may be even less prominent in this context than for other types of policy games. Indeed, they are hardly mentioned in writings on policy exercises. Better understanding of other's values and arguments features not as an important objective per se, but only as an intermediate step towards resolving substantial disagreements (cf. Brewer, 1986). And while Duinker et al. (1993, 3) cite 'foster[ing] communication and mutual understanding' as one of the three principal goals of their policy exercise on European forest policy, they do not follow up on this aspect in their extensive account of the exercise runs.

So what about our expectations concerning relational learning in the policy exercises that we designed and ran? Our interventions were quite short and expert participants did not come from one cohesive group. Therefore, we did not expect that we be would able to find evidence of changes in the relationships among participants, at least among the experts. Rather, we hypothesised that stakeholder perspectives and insights into the roles participants were playing would constitute the core of participants' relational learning. Learning about the key interests and motivations driving the countries represented in the policy exercise was also the aspect that we focused on primarily in our assessment. We expected this to be relevant especially for the student groups, given their lack of prior exposure to the topic of the policy exercise and its context, but anticipated – albeit lower – learning effects also for the experts, few of whom were policy-makers in real life. Finally, we did not see any grounds for assuming that the topic of the exercise – EU burden-sharing vs. REDD – would have an impact on the degree of relational learning. Therefore, in the following analysis, unlike in the previous two chapters on cognitive and normative learning, we mostly contrast the results for experts and student groups as a whole, without further distinguishing between participants in the burden-sharing and the REDD exercise. In terms of measurements, we relied on self-reported data in ex post surveys and interviews, given the difficulties of devising a reliable recorded measure of relational change. Regarding the structure of this chapter, the next two sections examine our findings regarding participants' learning about other's perspectives and mind sets, focusing in particular on insights into the roles that they or others played during the exercise. This is followed by a brief discussion of other aspects which came

forward in the evaluation of the four runs and which, while they may not qualify as relational learning as such, nonetheless relate to the interactive, social dimension of the policy exercise method.

7.2 Learning about perspectives across the four runs

7.2.1 Learning about perspectives in the expert groups

The experts taking part in the policy exercises acknowledged the relational learning potential offered by the role-playing aspect of this appraisal method. As one participant in the burden-sharing exercise reflected based on his own experience,

> *'when people are forced to play roles and are faced with the practicalities from a point of view of different countries or different individuals or different institutions this is a very powerful way of conveying information in a way that allows people to understand some of the difficulties that different countries might have.'*

Figure 15 summarises the results for the question on this dimension of relational learning (increased understanding of other stakeholders/actors as a consequence of the exercise) across three participant groups.[38]

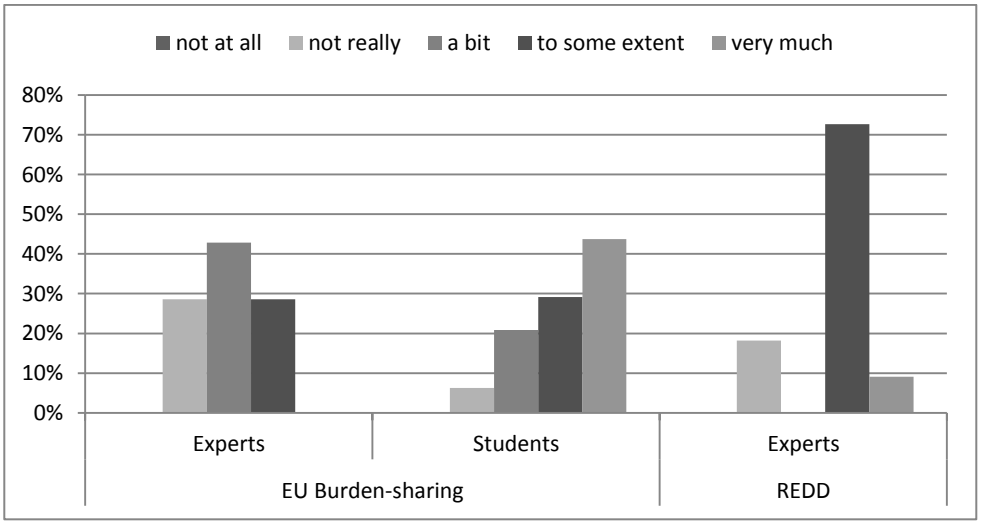

Figure 15. Self-reported relational learning (improved understanding of perspectives of other stakeholders).

[38] Unfortunately, there is no data available for the REDD student group as the block of self-reported learning questions was mistakenly omitted from their ex-post questionnaire.

Looking at the survey results, among the experts taking part in the burden-sharing exercise, opinions varied, with four participants reporting that their understanding had 'not really' increased, against ten who responded 'a bit' or 'to some extent' (mean 3.0). Contrary to our own expectations, the picture was quite different for the experts in the REDD exercise, where ten persons stated that their understanding had increased 'to some extent or 'very much', against only two who answered 'not really' (mean 3.73). Overall, we can thus conclude that the policy exercises did lead to some extent of relational learning in both expert groups. Testing for between-group differences using analysis of variances (ANOVA) revealed that the difference between the burden-sharing and the REDD expert group was statistically significant ($F(1,23) = 4.63$, $p < .05$). The higher level of self-reported relational learning on perspectives in the REDD expert group may be due to the fact that REDD is a relatively new addition to the agenda of the global climate negotiations. REDD financing is currently still being discussed in rather general terms, with the implication that REDD experts may not be as aware of countries' interests and priorities on specific issues, even where these have formed already, thus providing higher potential for relational learning through the exercise. On the whole, however, the difference in appreciation of the roles/perspectives dimension between the two groups was not reflected neither in the individual ex-post interviews nor in the 'three key insights' from the policy exercise that participants were asked to report in response to an open question in the post-survey.

Turning to data from the ex-post interviews, a first recurrent theme was the sense of having experienced 'first-hand' the complexity and challenges in navigating multilateral negotiations (especially when, as in the international climate negotiations that provided the context for the REDD exercise, decisions need to be taken by consensus) and in arriving at an outcome that is acceptable for all involved. In particular, the exercises appear to have made participants more aware of the (domestic) constraints and strong interests underpinning countries' positions even on seemingly technical aspects of a policy, which complicate reaching agreement. Somebody summarised:

> *'For the participant you get a real sense of how difficult these negotiations are and how you have to guess what kind of vested interest the other countries have. In my daily work I focus on environmental benefits and you don't try to understand so much why countries are against something but you just think the countries is [sic] not doing their job. It makes you realise what the agenda is of these countries and how these interests play together.'*

In a way, it seems surprising that these considerations should be novel to experts working on this issue area for a long time, albeit from different professional perspectives. Yet playing a role seems to have driven home these points more forcefully than may be possible in other contexts, as illustrated by the following quote:

> *'It [the policy exercise] has certainly helped me to understand some of the dynamics that we saw developing between those countries that were largely acquiescent to our request and those that weren't, which is very useful for the work that I'm doing at the moment.'*

These reflections were also echoed in general terms in some of the answers in both expert groups to the 'key insights' survey question, such as 'roles of different players', 'strategies of players', or 'players don't behave in a rational way.' Also in the debriefing to the REDD policy exercise, the insights gained from playing a role were the first aspect mentioned by participants in response to the question how it felt to take part in such an exercise.

The expert participants referred to the role dimension of relational learning not only in general terms, as described above, but also in terms of their specific experiences with or insights into certain roles. Most often, these observations related to the role that they themselves had assumed, but in some cases also concerned other roles represented in the exercise. For the EU burden-sharing exercise, it was the position and relative power of the European Commission that drew most attention. Asked whether playing the Commission role had furthered his insight in the way that actor operates in reality, one participant replied:

> *'Yes of course, I found this one of the most surprising things. When counter proposals were made in the negotiations, as a researcher, I might have raised all kinds of arguments against them. But since I was playing the Commission, I did not mention those points since I very much wanted an agreement to be reached. This happened very fast, by the way.'*

Another concurred that 'as the Commission, in the end, you can bring in very little, and you want most of all that an outcome is reached.' This view contrasts to some extent with one of the 'key insights' reported by someone else in the post-survey, who report having realised that the 'EU Commission has great framing power'. Ultimately, these two viewpoints may not necessarily need to be at odds, however. Recent analyses of European climate policy tend to agree that the Commission has substantially expanded its influence in climate policy over time (e.g. Jordan et al., 2010; Oberthür, 2011), with the swift adoption of the EU climate and energy package (which was still being negotiated at the time of the policy exercise) as a high-water mark. The Commission's successful framing of the set of individual policy proposals as a tightly integrated package that would ensure EU leadership in the international negotiations on the future of the climate regime was undoubtedly a key element to this success. Aside from the Commission's role, one participant also emphasised that the exercise 'made [him] realise what it means to have the presidency' in the European Union context.

In the REDD expert group, relational learning about specific roles seems to have been linked somewhat more to country groups than to individual countries. A first,

somewhat sobering insight was reported by a participant in the post-survey, namely that 'few countries actually have an incentive for an operational REDD mechanism'. If indeed applicable, this assessment would not bode well for the prospects of such a global mechanism ever coming about. As somebody elaborated in a follow-up interview:

> *'Already the technical complexity of REDD is mind-boggling, but the workshop made me realise that this is only part of the picture. In fact, politically speaking, only very few of the interests that the different countries have in REDD match and align, and in many cases they are outright contradictory. Especially the United States can keep pushing for highly cost-efficient and environmentally sound REDD in the negotiations, since what do they lose if it never gets adopted?'*

This reflection highlights the close link between the fate of a global REDD mechanism and progress in the international climate negotiations as a whole. While the initial idea of REDD held appeal for both developing and developed countries – promising additional funds to the former and supposedly quick and cheap emission abatement potential for the latter –, REDD in itself will not be a 'deal-maker' in the quest of reaching agreement on a future climate regime.

Another participant, who had represented a developing country during the policy exercise, reflected on the attitude of developed countries, the donors under a future REDD scheme:

> *'I got the feeling that donors are too demanding on delivery [of REDD credits]. I mean, at least in the beginning – and that may last for a long time – REDD will most of all be a learning process, and that in itself is beneficial. But if that understanding is not there [among donors], then very quickly REDD may become an elite activity for only some, very advanced REDD countries.'*

This line of thinking, expressing a differentiated view into the interests and mind set of REDD donor countries, echoes one of the 'key insights' reported on the ex post survey, namely that the 'first steps [need] to come from rich countries' in setting up an operational REDD mechanism. Finally, in two follow-up interviews, participants pondered the weight of Brazil in the international REDD debate. One interviewee felt that the exercise had illustrated to him that 'in essence, Brazil's motivation to tackle deforestation is domestic and what happens at the international level, [is] by definition less important' for the country. Another emphasised her key role as a 'bridge-builder between donors and developing nations' and as an 'example that with decent governance, you can stem deforestation pressures'.

Summarising, the qualitative data from post-interviews and -surveys provides evidence for some degree of relational learning among participants in terms of insights into the (constellations of) actors represented in the two policy exercises and

their motivations and interests. However, especially when comparing to the cognitive learning effects that we examined in the previous chapter, relational learning probably represents more of a side line of participants' reflections about the workshop. Some 'key insights' noted in the ex post surveys concerned this dimension, yet thoughts about actors and actor constellations were only offered upon specific prompts in the interviews, with a number of participants not having much reflection on this issue at all. In line with our expectations, we would thus argue that, at least for the expert exercises, relational learning, where it happened, was more of an ancillary rather than a main effect of the policy exercise.

7.2.2 Learning about perspectives in the student groups

A look back at Figure 15 which reports on the survey results for the question on self-reported relational learning (understanding of perspectives of stakeholders) reveals a rather high degree of relational learning among the students that participated in the burden-sharing policy exercise. 21 students replied that the exercise had increased their understanding of stakeholders' perspectives 'very much' and 14 'to some extent', with only three students responding 'not really' (see Figure 15). Analysis of variance (ANOVA) indicates that the mean score for this group is significantly higher than for the experts taking part in either the burden-sharing or the REDD exercise ($F(1,72) = 10.96$, $p < .05$). Our expectation to find a higher extent of relational learning among students than among experts is thus confirmed at least for the burden-sharing group.

On the basis of the qualitative data on students' relational learning about actor perspectives and roles, we find that while overall there may have been a higher degree of relational learning for this group, the patterns of what was learned are very similar to what we found among the expert groups.

As in the expert groups, a first relevant line of thought that emerged from the open question on 'key insights' in the post-surveys related not to insights into specific roles, but to the realisation of different national conditions and the complexity of interacting interests in international negotiations. As one person noted following the burden-sharing policy exercise: 'Amongst European countries there are different interests due to different economic development paths. This complicates the achievement of an agreement.' Others expressed similar feelings, emphasising the 'complexity of burden-sharing with countries in different stages of development (which also attribute different levels of importance to environmental policy in general)', that 'there's [are] so many different goals for all the different countries', that burden-sharing is 'a complex problem to manage as it includes a lot of actors and interests.' Reflecting on the correspondence between the simulation context in the policy exercise and the reality of policy-making in the EU, someone concluded that '[r]eaching an agreement between 6 countries [the number of countries that had

been represented in the policy exercise] is hard, 27 should be a nightmare!' Similar reflections ('all parties have different priorities', 'the different views and stakes countries have referring to REDD', '[t]here are different points of views and interests among developed and developing countries; and even among themselves.') were made, though somewhat less frequently, in interviews and surveys following the REDD policy exercise. One student put a particularly strong emphasis on this point:

> *'The most important thing I learned is that it is really hard for countries to come to an agreement. This is because countries are bound by their constituencies, their geo-political circumstances, by their history. And once you … play that role, you start to understand why there are so many arguments for and against and why there is conflict.'*

Another stressed the importance of the relational dimension for the quality of the agreement reached, noting that the 'outcome of negotiations also strongly hinges on mutual understanding.' In both groups, there was also some reflection on the relationship between individual and collective interest in negotiations: 'When it comes to negotiations on internal burden-sharing decisions, countries do not think a lot about the achievement of the aggregated EU target'; and in the context of REDD, '[t]he stakes of developing and developed countries are too different and they rarely seem to consider [the] general interest of the planet.' Some students in the REDD exercise also commented on the actor constellation with regard to specific issues, like this student who pondered the question of co-benefits, i.e. ensuring that the implementation of REDD does not focus exclusively on the carbon integrity of the project at hand, but also aims at wider social and environmental sustainability:

> *'Every country has a different stake when it comes to loss of biodiversity or indigenous rights. All these co-benefits mean different things for each country. So the more you include co-benefits the more difficult it becomes to reach an agreement.'*

Besides these more general relational considerations regarding the constellation of actors and their respective interests in the two policy exercises, a second set of relational aspects emerging from the post-interviews and -surveys again concerned students' perceptions and insights regarding specific actors that were represented in the policy exercise. Turning first to the exercise on EU burden-sharing, as for the expert run, it was the Commission's role that was singled out most often in this regard. One student, when asked what was the first thing that came to his mind about the policy exercise about two weeks after the run, responded the following:

> *'How frustrating it was to be the Commission. You think you are setting the agenda, work hard to consult all the member states and present a proposal that is a good, ambitious compromise. And then you are simply pushed aside at the decision stage.'*

This theme was picked up in a large number of responses to the open survey question. While some students mentioned it in general terms ('Role of the Commission'; '[t]he impact of EU Commission's [sic] attitude on policy-making', etc.), most of the answers provided a negative judgment of the Commission's influence in line with the interview quote above. Answers emphasised the '[w]eak power of the Commission in determining the crucial issues' and '[t]he difficult role of the Commission, in between the interests of all member states'. Apparently based on his or her own experience during the exercise, someone concluded that the 'role played by the European Commission is marginal. The CoM [Council of Ministers] changed dramatically the European Commission proposal' and another stated that the 'Commission really only presents its proposal, but in the end it's the member states who have the final word and their interests are very important.'

As in the expert group, however, there were also dissenting voices. One student observed that '[s]ometimes you cannot go for your own objective because it does not follow the Commission who finally has a lot of power in decisions making.' Another, while abstaining from evaluating its impact, highlighted the '[i]mportance of a strong role of Commission. They need to lauch [sic] the debate with a well prepared proposal.' Besides the interest in the Commission's role in policy-making, the only other actor that was mentioned explicitly was the rotating EU presidency, with one student underscoring 'the importance of the … EU presidential nation.' Another reflected on the Council of Ministers as a collective, noting that for this body, 'it is difficult to adopt a long term view. The Country' interest [sic] prevails when choosing targets and measures.'

Relational aspects covered by the students that participated in the REDD exercise were more diverse in outlook and theme. While some students we interviewed after the exercise stated that insights into roles and actors had not been at the core of their learning experience, others, like this student, were more positive: 'When you read articles, you learn mostly about policies. The exercise gave the whole idea and the details of how REDD works, what are the different positions on REDD … This was interesting to see.' Another concurred, adding that

> *'…it was interesting to see the variety of stakeholders involved in REDD. I didn't realise that so many countries have interest in REDD and would actually be willing to put money into it. I also learned a lot about the positions of the EU and the US in particular.'*

On the subject of the United States, as in the expert exercise, somebody expressed a pessimistic view of the country's commitment to REDD: 'I find it hard to believe that the US will participate meaningfully in the near future. The United States has no interest in taking part in REDD,' whereas Brazil was thought to have a 'crucial position in REDD'. Other comments in the post-surveys emphasised the divide between potential REDD donors and host countries on the one hand ('North-South

differences'), or, within the Southern group, the 'large differences between the LDCs [least developing countries] and the developing [emerging] economies, such as Brazil'. These considerations led one student to reflect on the perceived inequality in bargaining power in international negotiations: 'Interests of developing countries will always get less priority. It was striking how even with an exercise in which no party has real power, developing nations still end up with a bad deal for them.' Concerned with the same issue, another called on the rich countries to 'see also the benefit in … [the] long run in this issue, this would balance the negotiations.'

In summary, whereas references to learning about actor perspectives were on the whole somewhat more frequent and emphatic among students compared to the experts – an impression which was confirmed at least for the burden-sharing group by a significantly higher score on self-reported learning on this matter than the other groups –, what was learned in this regard seems to have been largely similar across runs. In all groups, relational learning about actor perspectives revolved around two main dimensions. Firstly, the exercise seems to have highlighted to participants the extent to which interests and positions in the policy arena, be it for REDD or for burden-sharing, are diverging, and what challenges this poses for reaching agreement. Secondly, there was relational learning about the interests and stakes of specific actors, which in the case of the burden-sharing simulations, focused largely on the Commission, but was somewhat diverse for the REDD exercise.

7.3 Other relational aspects

As set out at the beginning of this chapter, we largely concentrated on the dimension of actor perspectives and roles for our measurements in the relational domain. Since most expert participants did not know each other before the policy exercise and since the event lasted for only one day, other dimensions of relational learning (improved personal relationships, enhanced trust and better ability to cooperate) seemed less relevant for the specific context of our policy exercise. And while the students taking part in the policy exercise, who were studying in the same, rather large, MSc programme, were more familiar with each other beforehand, also here, we considered notions of trust and ability to cooperate as largely ancillary to the main learning goals of the exercise, which were more cognitive and normative in nature. What we briefly discuss in the following, however, are some aspects that, while not directly representing relational learning per se, do touch upon the social and communicative dimension of policy exercises as an appraisal and workshop format.

7.3.1 Expert runs

A key advantage that several experts emphasised when we asked them about their views on the strengths and weaknesses of the policy exercise method was its interactive character. As somebody commented, 'it was very involving, much more

than other workshops. Normally you at least have a chance to stay out of the discussions or give very little input but this really relied on everybody to give a lot of input.' Another participant agreed: 'everybody is stimulated to work together and to act, and this makes it a particularly active workshop form.' One participant pointed to the role-playing element of the exercise as a catalyst for communication among participants:

> *'It [the role-playing element] had a huge impact. Just the idea that everybody represents a country and not himself anymore and that you're supposed to behave like you're in real negotiations makes it really easy to talk to other people because you immediately have a topic to talk about, for example about what you think about some burden-sharing idea – I think it has created a spirit that was really full of respect and fun.'*

It thus seems as if the emphasis on interaction, negotiation, and exchange of arguments among participants had both substantial ('everybody is forced to share their expertise' in the words of yet another interviewee) and relational benefits. Somebody reflected after the REDD policy exercise:

> *'The post-workshop drinks are always a good indicator of how a workshop went. In this case, everybody stayed around and for a long time, which indicates how well we've connected and that people are interested in taking the networking further outside of the game context.'*

For the expert run of the REDD exercise, we had also included a specific question on the relational dimension to the ex-post survey. Participants' reaction to the statement 'the REDD-ALERT policy exercise helped to get a better grasp of who is involved in this issue and to build up relationship' was moderately positive. Only two participants (17 per cent) ticked the box 'not really', whereas all others were more sanguine about it, with two persons responding 'very much' and four each (34 per cent) 'a bit' and 'to some extent'. General satisfaction with the policy exercise was echoed in the unprompted reactions we received after the workshop. Emails from different participants thanked for the 'really enjoyable workshop' and the 'excellent meeting of yesterday'. The fact that one participant noted 'friends' as one of his 'key insights' on the ex post survey also seems to indicate that the expert exercises succeeded in social terms.

7.3.2 Student runs

The echo from students on the relational or social benefits of the policy exercise was also relatively positive. The majority of the students we interviewed following the runs stated that the policy exercise had somewhat increased their familiarity with their class mates. In most cases, this seemed to relate primarily to other members in their country team rather than the wider group of participants, as evident from the following quotes from three different post-interviews: 'I did get to know my study

mates. Some people in my group I didn't know before and afterwards I did. The exercise was also nice in that sense.' 'One of my team mates I didn't know before and now we talk to each other frequently.' 'I was in a group with three other people, who I didn't talk to before. The game helped me a lot with making some new contacts. It's a good game for that.' This view was not universally shared, however. On the question whether the exercise had helped her build new relationships with some of her class mates, one student replied: '[n]ot really. There was one girl in my group that I got to know a bit better … I wouldn't say that it was really suitable for getting to know someone.' Another student also declined that this had happened, arguing that '[w]e knew each other pretty well beforehand.'

In the ex-post survey, the students were asked how the policy exercise as a teaching method compared to other classes that they had attended during their studies, in Amsterdam and elsewhere. There was overwhelming agreement that it had been 'more interactive' in nature; 94 out of the 127 students (74 per cent) who completed the survey after either the burden-sharing or the REDD policy exercise agreed 'very much' with this statement, with only five students (four per cent) responding 'not at all' or 'not really'. Almost half (57 out of 123 respondents, or 46 per cent) also 'very much' agreed that the exercises had been 'more fun' than regular classes, a notion that only six students (five per cent) opposed. While aspects like of interactiveness and 'fun' clearly do not equate relational learning as we have defined it in Chapter 2, they do emphasise the 'soft' benefits of policy exercise as an occasional teaching tool in higher education. We will return to this issue in the following, final chapter.

7.4 Summary

Given the mostly short duration of the interventions and their focus on advancing knowledge, relational learning objectives are generally not at the core of the policy exercise method. One aspect of relational learning that did prove relevant in this context is their potential to improve understanding of other's mind sets and perspectives, in particular with regard to the roles represented in the exercise. Both experts and students reported insights related to the complexity of the international actor configuration at large, but also on the interests and constraints faced by specific actors in the policy context that was being simulated. While we had not expected the topic of the exercise to matter for participants' relational learning, it seems that for the expert group, the REDD exercise had greater effects in that regard than the one on EU burden-sharing. One possible explanation here again relates to the relative novelty of the topic which may have caught participants' less aware of the specific interests and agendas of the different countries concerning REDD financing. Finally, expert participants emphasised the value of the highly interactive character of the exercise, which seems to have facilitated communication and networking among the participants – a finding which while it does not equate relational learning as such,

nonetheless underscores the value of the policy exercise method also in the social domain.

8. Conclusions

Policy exercises have been proposed in the late 1980s as a way to make sense of complex environmental problems and learn about potential ways of addressing them. Since then, a number of these interventions have been implemented around the world, some of which have focused on climate-change related topics. Yet despite the significant investment needed in terms of time and resources to prepare such interventions, to date their potential – and their limitations – for stimulating learning benefits have been hardly evaluated in a systematic fashion. This is the gap that the present thesis seeks to address. The thesis demonstrates that policy exercises can produce substantial learning effects that can be empirically measured and documented. It developed and applied an innovative conceptualisation of learning effects that differentiates between cognitive, normative and relational learning. The results obtained illustrate the value of a typology that – rather than privileging one learning type over others – treats them on an equal footing. A tailor-made evaluation framework drew on a variety of instruments to measure each type of learning, allowing for capturing different aspects of each of them. The use of pre- and post-concept maps probably constituted the most important methodological innovation in this regard.

This concluding chapter first draws together the learning effects that we found across our policy exercises and discusses the implications of these findings for the policy exercise method (Sections 8.1 and 8.2). It then turns back to reflect on the learning typology and measurements and outlines a research agenda for both fields (8.3 and 8.4). The chapter ends with a research agenda (8.5) and some considerations on the recent changes in the landscape of international and European climate policy (8.6).

8.1 Learning effects from the policy exercises

The three preceding chapters analysed participants' cognitive, normative and relational learning from the policy exercises that we designed and ran. We found systematic empirical evidence of learning as measured by several mutually independent methods of observation. The evidence was stronger for individuals' cognitive than their normative or relational learning (Section 8.1.1). As expected, learning effects among students surpassed those among experts, and learning effects from the exercise on the more immature policy topic, REDD financing, were higher than those from the exercise on EU burden-sharing exercise, a more mature policy issue, in many regards (Section 8.1.2). Determining firm correlations of other, contextual factors with the degree of learning achieved proved more difficult (Section 8.1.3). Lessons for the organising team on how to design future interventions point to

trade-offs regarding their duration, the degree of structure and the amount of background information provided to participants (Section 8.1.4).

8.1.1 Cognitive, normative and relational learning effects across the four runs

Comparing the three types of learning over the different exercise runs, the evidence was strongest for participants' cognitive learning, which we had defined as encompassing the acquisition of new and the improved structuring of existing knowledge. Results from post-surveys and interviews indicate that participants gained insights both with regard to current policy developments and future policy options. More than anticipated given experts' experience in policy-making, policy consultancy or applied research, they also learned about the dynamics of policy-making and negotiations. This underscores the value of hands-on, experiential learning for developing a deep understanding of process dynamics – whether in climate policy or in a different field. Still in the realm of cognitive learning, analysis of pre- and post-concept maps testified to changes in the emphasis of key concepts associated with the topic of the policy exercises and to improved knowledge structuring. This was evident from an increase in the average number of map levels and items (the experts taking part in the burden-sharing exercise formed an exception here, but had started from a comparatively higher baseline, see Chapter 5).

The extent of normative learning – changes in participants' viewpoints, norms, or paradigms – that we could ascertain was by comparison more limited. When it came to objectively 'recorded' normative change, as measured through shifts in means in the assessment of a series of normative statements from pre- to post-measurement, we found only little evidence. There was moreover hardly any indication of normative convergence within the group, as expressed in decreases in the standard deviation of a group's survey responses from pre- to post-test. Participants self-evaluated the extent of their normative learning somewhat more positively, although rather in terms of revised viewpoints on specific issues than larger-scale normative shifts (see Chapter 6). The fact that participants (who primarily came from the global North) playing developing countries in the REDD exercise reported significantly more normative learning than those playing industrialised countries lends support to the hypothesis that role switches can support and facilitate normative learning, a finding which resonates with some views expressed in the literature (see Chapter 4 and Parson & Ward, 1998).

Finally, relational learning benefits from the exercises related primarily to a better understanding of others' mind sets, positions and preferences. More specifically, participants reported insights into the interests of countries represented in the policy exercises and their underlying interests and motivations, as well as the wider actor configuration (see Chapter 7).

8.1.2 Between-group differences in learning effects

We collected data on learning effects from two different policy exercises. The first focused on distributing emission reduction commitments in EU climate policy, a process commonly referred to as burden-sharing. The second examined options for designing a global fund for REDD, a mechanism under the global climate regime which aims at compensating developing countries for their efforts in reducing emissions from deforestation and forest degradation. We kept the design of the two exercises as similar as possible, in order to limit the 'noise' from additional factors that might lead to differences in the resulting learning benefits. For both topics, a first run was organised with a group of international experts and policy-makers, followed by runs with master students in Environment and Resource Management at the Vrije Universiteit Amsterdam. We used the student groups to collect additional measurements to gain further insights into the types of learning that an intervention like a policy exercise might trigger for this slightly different group of participants. This research design allowed for a two-way comparison, firstly between the impact of the two topics at the heart of the exercises that, as we argued in Chapter 3, differed in their maturity and their relative position in the policy cycle, and secondly between experts and students. Given the scarcity of previous research in this domain, we could only formulate tentative expectations regarding the impact of these differences on cognitive, normative, and relational learning. Still, it seemed plausible to expect that both the level of expertise and the maturity of the subject under investigation would impact on what participants might learn. Table 22 below summarises the expectations formulated in Chapter 3 together with the respective findings. The table zooms in on key results along the main axes for comparing learning effects set out in Chapter 3. Additional insights reported in the previous chapters, including on relevant conditioning factors and possible cross-correlations, are summarised in Section 8.1.3.

	Expected learning foci within the three learning types	**Expected differences between expert and student groups**	**Expected differences between the more mature policy issue (EU burden-sharing) and the less mature one (REDD financing)**	**Expected relevance of contextual factors**
Cognitive learning	Understanding/m awareness of policy options and their feasibility; both for current and future policy – CONFIRMED Policy dynamics (especially for students) - STRONGER THAN EXPECTED ALSO FOR EXPERTS	Higher for students than for experts – CONFIRMED	Higher for REDD financing (only for expert groups) – CONFIRMED FOR FUTURE POLICY	Higher for participants with lower baseline knowledge; for experts further removed from policy-making and students with less related academic backgrounds – CONFIRMED ONLY FOR EXPERTS FURTHER REMOVED FROM POLICY-MAKING
Normative learning	Changes in the perception of the desirability and feasibility of policy options and design features – CONFIRMED, THOUGH LIMITED OVERALL	Higher for students than for experts – CONFIRMED	Higher for REDD financing (only for expert groups) – CONFIRMED	Higher for participants with lower baseline knowledge; higher for roles further removed from reality – NO EFFECTS FOUND
Relational learning	Perspectives and interests of the countries represented in the policy exercise – CONFIRMED	Higher for students than for experts – CONFIRMED	No differences anticipated – REJECTED, HIGHER FOR REDD THAN FOR BURDEN-SHARING	No specific expectations – NO EFFECTS FOUND

Table 22. Learning effects – summary of expectations and findings.

The first general conclusion we are able to draw is that the effects for the student groups have been greater than those for the experts, at least where our measurements lent themselves for between-group comparisons. This emerges from the self-reported data (post-surveys and interviews) for all three learning types. It moreover shows in the recorded measures (concept map metrics) for cognitive learning. The higher learning effects we measured for the student groups were likely a function of their much less developed background knowledge on the policy context and the specific topic of the policy exercises. This lower knowledge base facilitated the assimilation of new knowledge and facts (cognitive learning), made the revision of pre-existing viewpoints (normative learning) more likely, and enabled novel insights into actor positions (relational learning). Yet whereas the learning effects were greater for students, we also found significant evidence of learning in the expert group. This is remarkable given the short duration of the policy exercise and the high baseline from which the experts started out, a point we will come back to in section 8.2.

What can we say about the impact of the exercise topic on learning effects? The comparison here is imperfect as we did not have the same experts participating in both groups (which, given the requirement of specific expertise and experience on the subject matter of the workshop, would have been impossible in any case). But we did use the same approach for identifying and inviting potential participants and are confident that the resulting mix of attendees and their level of expertise were largely similar in both cases. On the whole, our expectation that the greater 'novelty' of the issue of REDD financing compared to EU burden-sharing might enable higher cognitive and normative learning effects among experts (for students, both topics were equally unfamiliar) was confirmed. Drawing on interview and survey data, REDD experts self-reported significantly more normative learning, significantly more cognitive learning about future policy options, as well as more relational learning about actor interests and positions. There is thus considerable evidence that a more unstructured, less mature topic implies a higher potential for reaching specific learning objectives. Yet since the policy exercise drew on only two cases and a limited number of observations, this is only a tentative conclusion which merits further investigation.

8.1.3 Enabling and impeding factors for participants' learning

In addition to the comparing the learning affects along the main axes (level of expertise, topic of the exercises) discussed above and summarised in Table 22, we also probed into possible connections between certain characteristics of participants and their cognitive, normative and relational learning from the exercise runs. In general, we found only few robust links between factors like participants' backgrounds, their self-reported expertise and assigned roles during the policy exercise and their respective learning. A notable exception is the significant

correlation between 'softer' motivations to participate in the exercise - such as interest in the policy exercise method, in networking or in the project hosting the policy exercise (as opposed to noted interest in the topic of the two exercises) - and levels of self-reported cognitive and normative learning. We might thus hypothesise that an open orientation towards the method and its benefits seems to facilitate learning. This interesting correlation seems to call for selecting participants on their motivations if maximising individual cognitive and normative learning effects is the objective. Simultaneously, it also highlights the relatively greater challenge of stimulating learning in those stakeholders that are a priori less interested in engaging in policy games. Their learning may however be crucial to progress in the policy arena: our experience with preparing the policy exercises suggests that it is harder to convince (senior) policy-makers as opposed to other experts to participate. Be that as it may, our findings suggest that this subject deserves more comprehensive and systematic research into the relationship between exercise participants' motivations and learning results.

We were also interested in some broader aspects that participants perceived as conducive to or constraining of their individual learning experience and inquired about those in the debriefing sessions, post-surveys and interviews. Responding to questions in the post-survey prompting participants to evaluate a number of elements of the policy exercise (the exercise scenario, the role descriptions, facilitation, the debriefing session, etc.), all participant groups considered discussions with members of their *own* team as most useful to their individual learning experience. This assessment corresponds to our own observations during the morning sessions of the exercises (country-internal deliberations), where we witnessed participants quickly engaging in lively discussion with their team mates to assess the various policy options and their implications for the country they were representing. In our view, the way in which especially the experts used this setting to intuitively draw on and share their own expertise and experience underscores the communicative and deliberative value of policy exercises as 'boundary objects' between stakeholders coming from a range of different backgrounds (see next section). It also retroactively validates a number of decisions we took in designing the policy exercise, building on earlier experiences by Parson (1995) and Wenzler (1993). For one, in terms of steps of play, it seems to have paid off to start the day in a more cooperative, deliberative set-up (country teams developing their own strategy) instead of immediately moving into a more confrontational and rhetoric-driven negotiation mode. Equally, our decision to form country teams instead of assigning individual roles to participants (which would have had the benefit of being able to represent a wider range of countries) helped to stimulate in-depth discussion of policy content and to avoid an exclusive focus on antagonistic back and forth - between countries.

The element that participants on average considered least useful for their learning experience were the exercise scenarios (except for the students taking part in the REDD exercise, who awarded slightly higher scores). This feedback relates to a point that especially the experts brought up in the debriefing sessions and ex-post interviews, namely the appropriate level of background information about the 'state of the world' and their own country in the fictional future scenario and the role descriptions (see next section). The ex-post surveys also included a range of statements about aspects that may have impeded respondents' learning, such as an ill-chosen topic, a too short duration of the exercise, constraining scenarios or roles, over-structuring of discussions through the policy element cards, lack of knowledge or diversity among participants. The resulting mean scores per group were in general quite low (mostly in the range 1.5-2.5 on a 5-point Likert scale).This indicates that most participants apparently did not perceive these aspects to have been serious impediments. Among the experts, the short duration of the exercise (one day) received the on-average highest level of agreement (3.13 for the burden-sharing group, 2. 67 for the REDD financing group). The over-structuring of the exercise through the policy element cards also scored comparatively high among the burden-sharing participants (mean 2.97), but less so for the REDD experts. We will come back to these issues in the following section. In the two student groups, lack of knowledge among the participants received the highest mean score (2.59 and 3.17 for burden-sharing and REDD, respectively). This was not surprising, given that the rather complex subject matter of the exercise had been designed primarily with the expert audiences in mind.

8.1.4 Lessons learnt by the organising team

In our measurements and in the foregoing analysis, we have largely focused on what our participants learned from the policy exercises. But what lessons do we as the organising team take away in terms of how to design this kind of interventions? In preparing the exercises, we took great care to build on previous experiences (Duinker et al., 1993; Geurts et al., 2007; Parson, 1995; Toth, 1988b; Wenzler, 1993) and wherever possible sought to heed the advice provided in the respective accounts. Nonetheless, our experience again shows that designing such interventions involves trade-offs; we will briefly discuss the most important of them below.

The first trade-off concerns the ideal length of a policy exercise. Previous events of this kind have lasted from a couple of hours to up to three days. We opted for the middle ground by limiting the actual proceedings to one full working day. However, we started with a workshop dinner on the eve of the exercise to facilitate participants getting to know each other, and to shorten the preliminaries on the workshop day. In retrospect, we probably made the right choice in this regard, although several expert participants criticised the short duration of the exercise and expressed the view that the final agreements reached had been somewhat 'rushed'. Adding another half or

full day to the workshop would have allowed for more in-depth deliberations and exploration of some key issues. Yet it would have risked deterring some experts from taking part at all (as was expressed by interviewees during the ex-post interviews). In our view, a longer duration is likely worthwhile if the organiser of the event has sufficient convening power and/or influence over participants' agendas (i.e. if the exercise is commissioned by a higher level of hierarchy within an organisation, as is often the case in a corporate environment). However, this needs to be weighed carefully against the objective of ensuring sufficient quality and diversity among participants: in our case, academics would probably have found it easier than bureaucrats to allocate several days to our exercise. This issue might be circumvented, however, if a policy exercise was convened by an institution charged with policy-making, in which case its employees would be expected to attend rather than having to justify the absence from their desks.

Related to the first trade-off, more time for the exercise would also allow for working with two (or even more) consecutive scenarios as frames for the proceedings. In the burden-sharing exercise, two groups went through the same steps of play in parallel, but under different scenarios. This attempt did not prove very successful, however, and we therefore refrained from repeating the two-scenario set-up for the REDD exercise (see Chapter 4). It is possible that the scenarios were not sufficiently different or simply not forceful enough to leave a real imprint on the proceedings, or that the debriefing did not dedicate sufficient attention to this aspect. We would imagine that consecutively going through two different scenarios might prove more fruitful, but this would require a longer timeframe for the workshop.

Finally, there are two broader aspects that future organisers should carefully consider when preparing their own interventions of this kind, relating to the degree of structure and background information in a policy exercise. We had opted for a relatively structured format but only a limited amount of information through the scenario and role descriptions; in retrospect, we would probably revisit the former decision, but not the latter. We pre-structured the policy options at the core of the exercise by means of 'policy element cards' which country teams had at their disposal. These outlined alternatives for various policy design aspects of the policy decision that participants were confronted with. We had aimed to thereby facilitate participants' start into the exercise and to provide them with initial 'food for thought'. Looking back, specifying the different parts of the policy proposal to be negotiated definitely helped organise and focus the deliberations. Yet the amount of cards to be considered and the requirement to take a decision on each of the design features may have prevented an even more in-depth, creative discussion of the core aspects. As one of the experts who took part in the burden-sharing exercise reflected afterwards,

> *'[t]his was meant to try to replicate the real life negotiations on burden-sharing in 2020. This is only useful if more freedom is given to the actors to come up*

> *with ideas of what can be important in burden-sharing … I think this maybe didn't allow for a possibility to think out of the box.'*

Based on survey responses and ex-post interviews, the policy element cards were received more positively in the REDD exercise. This may also be due to the less mature nature of the topic of REDD financing, where a certain degree of pre-structuring of the discussion through the cards drawn up by the organising team may have been perceived as more useful and necessary. Yet also in this case, a focus on fewer core elements of the required decision and a smaller set of alternative options for each element might in retrospect have been advisable.

A second question relates to the appropriate level of background information for the exercise through scenarios and role descriptions. During the debriefing, expert participants of the burden-sharing exercise discussed at length the pros and cons of providing more data or real-time modelling support. Possibly, more specific assumptions on countries' emission trajectories, abatement costs and potentials, as well as model results that would have provided information on the consequences of key policy options being considered might have improved the quality of deliberations and the outcomes of the exercise. Participants remained divided on this point. Whereas according to one person in the ex-post interview, 'simple evaluation tools would have added even more bite to the discussions', others disagreed:

> *'My first thought during the day was that we did not have enough information, but on reflection I thought that more information might not give you a better understanding of the issues that are faced by people, because political decision making is so non-linear; it is not illogical as such, but it doesn't follow a logical linear format. Therefore simply giving more information may well make things more difficult instead of enlightening anyone.'*

The last point made in the quote above is indeed the crux of the matter. Arguably, more (quantitative) background information might give participants a better base for decision-making. This in turn could enhance the realism of the exercise and potentially improve the relevance of its outcomes. At the same time, such an approach carries a number of risks. First, the production of such data necessarily involves a large set of assumptions. What these are based on and to what extent they seem credible to participants could become a serious subject of contention and thus risk side-tracking deliberations from the overall objectives of the policy exercise. Second, real-time modelling support would mean adding a level of complexity that would likely overwhelm participants of a short intervention or at least detract their attention from its primary goals. Such a workshop design would then rather serve the goal of field-testing the model and not the development and testing of options for future policy. Third, the question remains how the information gained from the model outputs would be used in the deliberations. Parson (1995), whose policy exercise offered participants access to an integrated assessment model, reports that

while country teams initially did not make use of the model at all, they later did so primarily with partisan intent (finding additional arguments for the positions they had already developed) rather than a genuine desire to assess the implications of certain options. In conclusion, there are good reasons to consider building in a stronger quantitative element into a policy exercise. Yet we would argue that it may be primarily worthwhile for a larger-scale intervention, where sufficient capacity is available to develop or adjust an existing model to the specific needs and goals of an intervention and where participants are involved over a longer timeframe, and ideally already during the preparation stage of the exercise.

Having thus summarised the proximate learning effects from our policy exercises we will now turn to the broader question what our findings imply for the value of policy exercises in climate policy appraisal.

8.2 The value of policy exercises for climate policy appraisal

Overall, the learning effects that we found for our participants resonate with the emphasis on cognitive learning benefits voiced in the theoretical literature on policy exercises. Policy exercises are, generally speaking, part of the rationalist appraisal tradition. Their primarily cognitive motivation is visible in the assumption that simulating policy interactions in a 'simplified model of reality' can yield insights into potentially promising and feasible policy options or ways in which institutions or decision-making processes might be reformed (see Section 2.2.3). In this sense, the predominance of cognitive as opposed to normative or relational learning effects that we have found for our own policy exercises is not an unexpected result. The extent of this learning was quite significant even for the expert groups, given the background knowledge that the experts already had as well as the short duration of our intervention. There was evidence for improved knowledge integration and structuring, and participants self-reported insights into the advantages and drawbacks of various policy options. A broader, more holistic view of the problem also transpired from some post-exercise survey responses and interviews, but this is not a finding we can generalise for the wider group of participants.

Indications of an enhanced long-term orientation and a serious reflection of the risks and contingencies involved in future policy-making were more ambiguous. While participants mostly refuted the notion that the exercises had succeeded in projecting them into the future scenario and stated that they had taken many of their decisions in the exercise by reasoning from the present-day situation, some of the evaluation data and the exercise outcomes do highlight risks and contingencies specific to *future* policy-making. Achieving a stronger 'futures orientation' may also be a question of exercise design (for instance, working with several consecutive scenarios may reinforce this goal, see above). Yet achieving this objective is known to be notoriously

difficult also when using other interactive appraisal methods, such as backcasting (van de Kerkhof, 2004). We unfortunately cannot claim that the policy exercise method, at least in the way we applied it, fared much better in that regard.

Overall, participants reported – and this corresponded with our own observations – that the exercises had succeeded in creating a very interactive atmosphere that had stimulated them to share their knowledge and engage in a lively dialogue over the proposed policy options and their broader impacts and associated challenges. Initially, this happened primarily inside, but later also across the different country teams. The 'decision pressure' created through the policy exercise format, with country teams having to formulate their position on different aspects of the policy problem at the heart of the exercises under time pressure, played an important role in fostering such an effective exchange of information across disciplinary and professional boundaries. It was striking to see how participants brought to bear their realms of expertise on the questions they were confronted with and made an effort to communicate this knowledge to others from different backgrounds. Therefore, like others (Parson, 1996b; Toth & Hizsnyik, 2008), we believe that there is significant promise in using policy exercises as a tool to stimulate cognitive and relational learning at the interface of science and policy-making. Our exercises exposed experts from academia and think tanks to the realities and dynamics of policy-making and negotiations (see Chapter 5) and helped them identify knowledge gaps and potential needs for future research. Government officials in turn learned about recent scientific findings relevant to policy questions that they may have had little time to consider in more depth in their day-to-day work.

Aside from individual learning benefits, to what extent could we observe that the policy exercises generated insights that could be considered novel and relevant at a broader scale, constituting a genuine gain in knowledge? As discussed in Chapter 5, the policy innovation we observed was fairly limited. The EU burden-sharing exercise went beyond current political discussions on how to distribute future emission reduction commitments on two counts: by developing a composite criterion on which future burden-sharing could be based, and with its suggestion to factor future adaptation needs into arrangements for mitigation burden-sharing. The REDD policy exercise on its part provided a 'test drive' for specific options that had been suggested in the academic and grey literature, but that had not yet entered policy discussions. The latter may point to a valuable function of the policy exercise method more generally: the reasons for which priori 'value-neutral' policy options like an auctioning design for a global REDD fund might be unpalatable to developing countries may most easily be understood in an interactive, future-oriented setting like a policy exercise. Parson (1995, 18) also comments that while his exercise had hardly yielded new policy options per se, a 'more interesting service of the exercise was to clarify and focus arguments for and against particular policy proposals.' Given that a large part of assessment and evaluation activities in environmental policy

focuses on technical learning on the programmatic aspects of policy (Huitema et al., 2011), this makes for a rather important niche for simulation-gaming approaches such as policy exercises.

In terms of generating new policy options, however, our results fall short of the claims made in the theoretical literature (Geurts et al., 2007; Parson, 1996b; Underwood & Duke, 1987). At the same time, they resonate with other accounts of previous policy exercise efforts. For instance, Duinker et al. (1993, 19) write:

> *'We had hoped that our policy-exercise workshop trials … would launch us well along the road to discovering some incisive policy insights in relation to the problems posed by continued forest decline in Europe. After all, one of the highest objectives of the policy exercise in theory is to learn policy lessons that would probably not be learned through conventional means [...] However, our initial reactions to the substantive outcomes of the workshops were that there was 'nothing new under the sun', that we had seen before all the policy messages and perspectives tabled at the workshops.'*

Parson (1995, p. 18) reports a similar experience: 'Although one objective of the exercise was to support the generation of novel policy ideas, few of significant novelty were brought forward.' And Joldersma (2000, 83) notes in an ex post evaluation of an intervention on health care in the Netherlands that 'no new policy alternatives were created'.

Producing ´new´ knowledge is a rather demanding yardstick for any one-off interactive appraisal. In fact, the very nature of policy games might potentially constrain thinking 'out of the box' and developing truly innovative policy options. In view of the structured character of these interventions, one should moreover not underestimate the role that the mental models of the game designers may play in shaping their outcomes as well as what participants take away from them. And whereas policy exercises are generally less structured than other types of policy games (as they do not involve a predetermined reward or pay-off structure for participants' actions, see Chapter 2), the scenario, role descriptions and other background materials (as in our case the policy element cards) nonetheless frame and bound the 'possibility space' for decision-making. In general, participants are invited to contest and critique the assumptions made, which in itself implies considerable potential for learning both among participants and organisers. Yet the basic approach still involves a significant 'pre-structuring' of the topic being simulated by the organisers. We found traces of this when analysing the concept maps that participants drew up prior to and following the exercise runs (see Chapter 5). On the one hand, it is only natural and desirable that aspects that stood central in the deliberations during the exercise gained prominence on the post-maps. On the other hand, the extent to which this happened does make us wonder whether there is not a tension between this kind of set-up and the proclaimed goal of creativity and

innovativeness. Ultimately, it may do more justice to the potential of the method if one were to view the outputs from policy games and exercises less as an end in themselves than as one step in a longer appraisal process (discussed in the literature as 'design in the small' - i.e. the individual game - vs. 'design in the large' - i.e. the overall learning or change process, cf. Klabbers, 2009). This is also what Duinker et al. (1993, 19) conclude:

> *' […] perhaps such workshops are not to be expected to generate the policy lessons themselves, but rather are to provide the raw materials […] upon which a more firm policy analysis can be based. We see this as being more realistic in terms of expectations of what can be accomplished in meetings of this kind [...]'*

After these somewhat sobering conclusions, what factors might increase the learning potential of policy exercises for policy development? A couple of aspects come to mind here. First, as mentioned above, a longer duration of the exercises may allow for more in-depth reflection and for iterative experimentation with different policy options and framing conditions. Connected to this, there is much to be said for a collaborative game design process that involves key stakeholders from the beginning. Such an approach strengthens their commitment to the project and helps ensure that the game design is adequately approximates reality. Moreover, the design phase itself can serve as a powerful learning experience. Mendler de Suarez et al. (2012, 38) refer to it as 'co-design as a bridging process', where stakeholders learn about each other's mental models as they collaboratively identify game elements and mechanisms that approximate the key dynamics of the real-world system they are operating in. Last but not least, and at the risk of stating the obvious, the right set of participants is key to generating both plausible and innovative outcomes from the exercise (see Chapter 3; cf. Cuppen, 2009). It has been our experience that it is possible to convince senior, well-connected experts from the research and think tank worlds to take part in such a workshop. Yet we, like others, have found it much more difficult to persuade policy-makers to make space for travel and a full-day event in their busy agendas, especially if its direct usefulness for their day-to-day tasks is not immediately obvious (see Chapter 4; Zhou, 2014; Bachofen et al., 2013). Ultimately, all three suggestions above require a high level of convening power and sufficient resources from the organisers of such events. This further raises the threshold for using an already relatively time- and resource-intensive method. If these conditions are met, however, the attendant learning effects may well justify the effort.

Finally, some words on the value of the policy exercises for our second participant group, the Master's students in Environment and Resource Management. Our measurements yielded substantial amounts of cognitive, self-reported normative, and relational learning for this group, thus underscoring the benefits of policy games as a particularly active form of student learning. Our own observations equally confirmed their potential in an educational context that is frequently emphasised in the literature (e.g. Harteveld, 2012). The exercises stimulated a high degree of

involvement among the students, all the while leaving them in control and letting them experience the consequences of their actions first-hand. Knowledge about complex subjects was assimilated quickly (cognitive learning), yet there was sufficient room for the exercises to reveal student misconceptions, thus giving rise to normative learning. Perhaps most significant, however, were the insights students gained about the dynamics of the policy process, an eye-opener for those who had not been confronted with policy-making before.

8.3 Cognitive, normative and relational learning revisited

Apart from examining the potential of policy games and exercises for climate policy appraisal, a second key goal of this thesis was to make much-needed headway in measuring learning in an appraisal context, both in theoretical and in methodological terms. In Chapter 2, we explored various facets of learning processes and outcomes as they are described in the policy sciences and environmental studies literature. We derived three types of learning – cognitive, normative and relational –, which in our view capture key dimensions of learning in an appraisal context and which lend themselves to systematically assessing learning across different cases and social units of analysis. The first two types – cognitive and normative learning – feature prominently in the policy learning literature. The last one, relational learning, reflects the emphasis on interpersonal, social aspects of learning which transpires from research on learning in natural resource management, with its focus on small groups of actors that become familiar with one another and that (begin to) collaborate. Our reasons for diverging from dominant conceptualisations of learning in an environmental governance context were twofold. First, we wanted to steer clear of a 'levelled' understanding of learning, preferring to treat the different types of learning side-by-side in a non-hierarchical manner. The often visible preference for learning that involves 'deeper', normative change in the existing literature seemed questionable to us. Therefore, we thought it both more appropriate and more systematic from an empirical viewpoint to analyse learning effects in terms of their nature (cognitive, normative, or relational) rather than their perceived value. Secondly, by separating out relational learning, we aimed at emphasising an aspect that in our view is key in the context of interactive appraisals, yet that is often treated only implicitly in other typologies.

Having applied this conceptualisation for measuring learning across the four policy exercise runs, we have found it to be very workable. It was well suited to operationalisation through the set of instruments that we introduced in Chapter 3 and that we will briefly return to below. The typology lends itself to cross-case comparison, also when involving different evaluation teams. Distinguishing phenomena along the cognitive-normative-relational axis is intuitive. Moreover, the lack of a hierarchy between different levels of learning may reduce evaluator bias as

there is no pressure to detect 'higher order' effects, which may not be forthcoming, especially if the appraisal process is rather short. This typology has by now been applied several times already. Huitema et al. (2010) used it to analyse learning effects from a set of citizens' juries on water management in the Dutch part of the Rhine basin. Based on pre- and post-surveys, concept maps and interviews, they found substantial effects for all three types of learning among the members of the citizens' juries, but much less for the policy-makers who had commissioned them. Munaretto and Huitema (2012) applied the typology to examine water and environmental governance processes in the Venice lagoon. Relying on interviews, participatory observation and archive data, they reported that learning among stakeholders in the Venice lagoon was largely limited to cognitive learning within the established management practices, with normative and relational learning being an exception rather than the rule. Finally, Baird et al. (2014) investigated stakeholder learning from an adaptive co-management process on climate change adaptation in Ontario, Canada. Using concept maps, social network analysis and self-reported measures, they found robust evidence for cognitive and relational learning as a result of the process, with normative learning outcomes being more ambiguous. These rather diverse applications of our conceptualisation illustrate its usefulness across a range of contexts. Moreover, their results largely dovetail with the findings reported in this thesis insofar as the evidence for learning was stronger when it came to cognitive learning effects as opposed to relational and especially normative learning effects, and that this evidence was stronger for less expert participants.

As with other classifications and typologies, there is the inevitable question of the interdependence between the different dimensions or categories. The three types of learning that we identify are certainly interrelated. Yet they nonetheless display specific, distinct characteristics which can and should be assessed separately if one is to come to a more sophisticated understanding of the benefits of interactive policy appraisals. For instance, whereas a higher degree of normative learning may indeed stimulate a greater desire and/or ability for cognitive learning or vice versa, neither is a necessary precondition for the other. In other words, it is possible to absorb new facts without changing one's view on the subject at hand. At the same time, deliberation and persuasion may lead a participant to re-evaluate known arguments and thus change his or her viewpoint without having increased his or her knowledge base. While it seems intuitive to assume that cognitive, normative and relational learning can in principle reinforce one another, trade-offs between the different types seem equally imaginable. Levy (1994), for instance, posits that a high degree of cognitive complexity may in some cases impede normative learning. This is because more complex, and consequently less falsifiable cognitive structures may 'make it easier for actors to assimilate inconsistent information into their existing belief structures as exceptions and special cases' instead of revising their viewpoint (ibid, 295).

Our evaluation framework had not been designed so as to tease out the interrelationship between the three basic types of learning that we identified. Instead, we had been primarily interested in assessing the occurrence or otherwise of the different types of learning under varying circumstances. This was primarily due to the state of the literature, where systematic assessments across learning types had thus far been lacking. When examining the learning effects found in this thesis from the angle of interactions between the different learning types, the results are inconclusive. On the one hand, the fact that normative learning was more limited among the expert groups seems to support Levy's intuition reported above that 'increases in cognitive complexity may be a way of avoiding belief change rather than facilitating it' (ibid). The same is true for the finding that normative learning was more pronounced for the less mature policy field. On the other hand, the general correlation between the different types of learning, especially among the student groups, seems to indicate that the different types of learning do indeed reinforce one another. As a consequence, we may speculate that the relationship is non-linear, decreasing with the degree of expertise. There is consequently an interesting research agenda in seeking to tease out this relationship, e.g. by conducting consecutive exercise runs with students and studying how increasing expertise impacts on degrees of normative learning. In order to assess the relationship across the entire range of expertise however, similar consecutive exercises would have to be conducted with experts as well because there may be one or more points of inflection.

8.4 Measuring learning in appraisals – pitfalls and promises

After these considerations on our theoretical framework, we now move to our measurements on learning and the instruments we used to that effect. In order to adequately capture the effects from the policy exercises, we designed a comprehensive evaluation framework that relied on a range of mutually independent measures of observation. It combined recorded and self-reported, qualitative and quantitative data sources, and measurements were taken at multiple points in time. We are aware of few evaluation efforts of deliberative appraisals reported in the literature that use a similarly broad array of instruments, especially including pre- and post-comparisons (for exceptions, see e.g. Baird et al., 2014; Huitema et al., 2010). We hoped that this approach would allow us to capture different aspects of participants' learning and to compare certain findings across data sources. Reflecting on our experience, we conclude that the framework we used has met these expectations, despite some challenges that we encountered and that we will discuss in more detail below. A more recent effort by Baird et al. (2014) builds on its elements, expanding the toolbox with social network analysis, yet placing less emphasis on qualitative data.

The use of concept maps for assessing conceptual change as well as changes in knowledge structure arguably represents the most important innovation in our evaluation framework. Pre- and post-concept maps and the associated analytical methods for dissecting them, while well known in education research, prove a promising new addition to the evaluative apparatus in interactive appraisal contexts. Tracing how the positioning of concept categories on participants´ maps shifted from pre- to post-measurement enabled valuable insights into the changes in knowledge structure as a consequence of participating in the policy exercise. This usefully complemented the self-reported data on key insights gained from the workshops. Pre-post comparison of other metrics, like the number of items and levels per map and especially changes in response types (lists, labels, explications) shed additional light on the question whether the observed shifts were also accompanied by an actual improvement of participants' mental maps of the issue area as reflected in the concept maps.

Although our measuring framework advances on the state of the literature, there are at least two avenues by which future research could further improve it. First, coding and subsequent data preparation for analysis are both time-consuming and potentially error-prone. Therefore, future efforts using this kind of data might focus on ways to computerise and/or digitalise parts of the process (see e.g. Cline et al., 2010), although manual coding will likely still be required at least to some extent. Second, in this research, we used the concept maps purely for assessing cognitive learning. In the future, it might be worthwhile to explore to what extent the first type of analysis (tracing concept centrality and specificity) might also allow for inferences about normative change. Conceivably, greater centrality of certain concepts and higher levels of elaboration might also indicate changing normative priorities.

A second key instrument in our evaluation framework was the pre- and post-surveys completed by participants that we used to gather data on both recorded (normative) and self-reported learning. Pre-test-post-test measures that seek to detect shifts in knowledge and/or viewpoints by asking participants to respond to a set of closed statements before and after an intervention are commonly employed in evaluation research, but less often in environmental appraisals (for exceptions, see e.g. Deyle & Schively Slotterback, 2009; Garmendia & Stagl, 2010; Huitema et al., 2010; Schively, 2007). While they are attractive from a research design perspective due to their quasi-experimental character, their implementation in a context like ours is not without challenges. Pre-post measures aimed at tracing increases in participants' knowledge (which would amount to recorded cognitive learning) are, given their quiz-like nature, hardly appropriate for expert participants. For this reason, we did not include them in our measurements. At the same time, formulating propositions ex ante that capture the essence of what might be learned in normative terms during a deliberative appraisal is notoriously difficult (Deyle & Schively Slotterback, 2009; Garmendia & Stagl, 2010). If evaluators wrongly anticipate the core questions around

which normative learning among participants revolves during the appraisal, they risk missing important learning effects. This underscores the value of triangulating such recorded pre-test-post-test data with other, self-reported sources that allow participants to share their own reflections on the process. Yet as Patton (1990, 467) points out, 'triangulation of data sources will seldom lead to a single, totally consistent picture', but may instead shed light on different aspects of the same phenomenon.

On the whole, we have made good experiences combining recorded data (concept maps for cognitive learning, pre- and post-survey statements for normative learning, as well as observations and the recordings of the exercise proceedings) and self-reported data (from surveys and post-interviews). Recorded data pointed to patterns that participants may not necessarily have been conscious of themselves, and which self-reported, qualitative data could sometimes help to explain, providing a level of reflection not commonly found in the former. Future research may look into devising similarly complementary recorded measures for assessing relational learning – social network analysis is one, but certainly not the only tool that comes to mind here. In conclusion, we agree with Blackstock et al. (2007) that future evaluation efforts in this field need to move from the purely anecdotal and qualitative that dominates this literature towards more systematic, ideally multiple-case assessments. And yet there remains an important role also for qualitative methods. Learning from appraisals is a process that is likely too complex and multifaceted to be satisfactorily captured by purely quantitative indicators.

This brings us to a broader point. As we discussed above, especially for the expert groups, the learning effects from the policy exercises that we measured were, while not insignificant, on the whole rather modest, certainly compared to some of the claims voiced in the early literature on policy exercises. We find the same pattern also with regard to other deliberative appraisals. Their alleged benefits are extensively covered in more theoretical work, and confirmed to some extent in analyses relying primarily on qualitative, anecdotal information. Yet in the few cases where attempts are made to assess outcomes more systematically, and where these also include longitudinal (pre-post) measurements, the record is often more mixed (see Chapter 3). Above we touched upon some of the reasons why more systematic measurements may not pick up on learning that occurred during the appraisal. This may be related to the difficulty of posing the right questions ex ante, the hesitation of knowledgeable participants to admit to having learned or having changed viewpoints during the appraisal, or the fact that some effects may only be realised at a later stage.

On the other hand, it is also possible that the link between deliberative appraisals and learning is more tenuous than is often assumed. As Garmendia and Stagl (2010, 1718) conclude based on their attempt to measure learning from three participatory

processes in natural resource management in Europe, 'we find that social learning does happen in participatory workshops, but … to a lesser extent than expected.' Their hypotheses for expected learning outcomes had included problem reframing, better understanding of others' perceptions and needs, enhanced capacity to deal with the complexities and uncertainties inherent in environmental change, also with a view to taking joint action – a fairly typical list for the public participation literature. It is still early days for such conclusions to be drawn at a general level, yet it is clear that more research, relying on robust evaluation efforts, is needed to examine to what extent the outcomes of deliberative appraisals in environmental governance justify the time and effort spent on them and what other intervening variables (type of participants, workshop design, timeframe; cf. Cuppen, 2009; Garmendia and Stagl, 2010) co-determine their impacts. In particular, more longitudinal, multiple-case assessments would go some way towards addressing these issues, by shedding light on potential longer-term effects and identifying key drivers and obstacles to success and failure.

As our experience as well as that of others shows, the challenges to a more systematic evaluation of learning effects from deliberative appraisals are not insignificant. Longitudinal assessments, for one, face the difficulty of separating the actual impact of the appraisal process from the 'noise' of other encounters and experiences that participants have gone through in the meantime, thereby reducing their potential usefulness especially for short interventions. Moreover, many appraisals involve only a relatively small number of participants. This limits the potential for quantitative data analysis. Appraisal processes often require a lot from their participants in terms of time, commitment and active input. It is therefore only natural that elaborate evaluation efforts risk being perceived as cumbersome additional burdens and possibly being avoided as such. This risk is perhaps even greater for pre-post measures that may seem repetitive and all-too academic to participants. Carefully building evaluation elements into the appraisal process rather than presenting them as an addition to it may provide a possible remedy here. Yet despite all difficulties, putting more effort into systematically measuring learning outcomes from interactive appraisals is indispensable if the field is to mature beyond the theoretical discussions, case studies and typologies that dominate the literature to date.

8.5 Towards a research agenda

The preceding sections set out various aspects of a research agenda resulting from this thesis in greater detail, in the context of our discussion of the respective findings. At this point, suffice it to summarise three key issues. First, there is still considerable scope in refining methods of measuring learning. While this thesis broke new ground in terms of its efforts to measure learning systematically, its success in demonstrating

that normative learning in particular took place was limited. Here, new proxies for such learning and methods for measuring it are called for, perhaps linked to the analysis of concept maps. It is, however, also conceivable that normative learning as a concept may have less traction than anticipated, at least when it comes to policy appraisals among experts. Were such a finding to harden in the light of further efforts to substantiate instances of normative learning in policy appraisals, this would have considerable consequences for our assumptions about appraisals and negotiations not just in the climate policy domain but beyond, with respect to policy-making and international relations (theory) in general.

Improving the measurement of learning is closely related to the second item on the research agenda: a better understanding of how the different facets of learning empirically relate to each other. In order to design better targeted interventions, we need to understand the conditions under which different learning types reinforce or compete with each other. Ideally, this would build on progress in the realm of measurement, but in reality pursuing these two objectives will likely be interdependent, with progress in one area enabling advances in the other.

The first two items on the research agenda refer to challenges that are likely amenable to incremental progress because both rely on tweaking established instruments and procedures. The last item is different insofar as there are no proven success stories yet. When it comes to unleashing the supposed creative potential of policy exercises for identifying new policy pathways, we need a more fundamental game changer. Can policy exercises really live up to their most important function, and if so, under what circumstances? In many ways this represents the 'holy grail' of the profession, the aspect that is simultaneously most important and least advanced. Given this status, it is perhaps not surprising that this thesis cannot claim to have made a breakthrough on this front. Yet the fact that it represents the most rigorous, systematic evaluation of policy exercises to date suggests that the onus now lies with proponents of the method to prove said benefits. We hope in any case that our work on the theoretical and empirical challenges and opportunities for learning will in various ways make a contribution to progress in environmental governance.

After these reflections on the theoretical and methodological contributions of the thesis and the resulting research agenda, we will return to the subject matter that our policy exercises addressed in the following final section: international policy efforts in the fight against climate change, and how they changed since work on this thesis started.

8.6 The changing nature of international climate governance

Our objective with this thesis was to contribute to a better understanding of the value of policy exercises for climate policy appraisal. The two policy exercises that we used to investigate this question focused on topical issues for the medium-term future of European and international climate policy. The first examined the dilemma of how to distribute emission reduction commitments among EU member states, an exercise known as burden-sharing. The second focused on design options for a global fund to compensate developing countries for reducing their emissions from deforestation and forest degradation (REDD). Both are subjects that are inherently distributive in character (see Chapter 3); both also emanate from the assumption of the continuation of a centralised, top-down policy architecture, be it at the international (in the case of REDD) or at the EU level (for burden-sharing).

Yet looking at recent developments, the landscape of climate policy appears to be evolving into a different direction, globally, but possibly also within Europe. Recent UN climate conferences have partly healed the wounds the 'trauma' of the Copenhagen summit (Haug & Berkhout, 2010) had inflicted on the multilateral negotiation process. The international community has embarked on a 'road to Paris' where a new, post-2020 climate agreement is to be negotiated at the end of 2015 (Grubb, 2013; Rajamani, 2012). Yet agreement on key aspects such as emission reduction commitments and finance in a future global climate accord remains elusive at this point in time. While it is likely that the UNFCCC will remain a key forum for climate change issues, future climate action looks set to be dominated by a pluralist, bottom-up patchwork of public, transnational and private efforts, initiatives, and actor coalitions at multiple levels (Stavins et al., 2014; van Asselt & Zelli, 2013). This would also have implications for the further development of REDD. There have been some advances on this issue at the 2013 COP in Warsaw, endorsing a fund-based rather than a market-based architecture for REDD (UNFCCC, 2013). Yet the question of financing sources for REDD remains unresolved and REDD remains but one 'piece in the puzzle' in the struggle to reach a comprehensive agreement. It is therefore likely that progress on the issue will mostly 'occur through complex fragmented pathways of international assistance, bilateral and multilateral agreements, and civil society and market-based processes' (Agrawal et al., 2011, 374) rather than a top-down process. In the meantime, while the provision of the sustainable long-term financing needed for REDD is certainly less secure without a global framework, REDD has started to transform the thinking of many developing country governments about forest issues and produced first impacts on the ground (Haug & Gupta, 2013; Sunderlin & Atmadja, 2009). Many of the aspects raised during our policy exercise, such as the verification of emission reductions from deforestation, remain relevant for performance-based REDD payments also in smaller-n or bilateral constellations, such as the carbon funds run by international development banks or donors governments.

European climate policy is equally confronted with new challenges. Over the past two decades, climate change has become an increasingly high-profile issue in the EU and has given rise to an impressive level of policy activity (Haug et al., 2010; Huitema et al., 2011; Jordan et al., 2012). The adoption of the 20-20-20 by 2020 targets and the comprehensive Climate and Energy Package in 2008 probably constituted the high point in EU climate policy-making to date (Jordan et al., 2012; Oberthür & Pallemaerts, 2010). Since then, the economic downturn and particularly the debt crisis in the Southern member states have pushed climate change more to the background of the EU agenda, severely reducing prospects for ambitious policy outputs in the short to medium term. The draft '2030 framework for climate and energy policies' for the post-2020 period released by the European Commission in January 2014 (European Commission, 2014) is a testament to this diminished ambition. However, despite this changed outlook, the Commission's proposal foresees that the fundamental tenets of EU climate policy be retained also over the next decade, including the distribution of mitigation commitments across Member States in the 'effort-sharing domain'. This once more underscores how deeply the notion of burden-sharing is engrained in the way climate policy is made in the EU. Jordan et al. (2012, 263) call the fact that 'policy harmonisation' in the EU – the adoption of collective policies and targets – has often only become possible through differentiated implementation one of the foundational paradoxes in EU climate policy.

Finally, there is another shift that has been shaping the climate policy landscape in recent years. With the failure to reverse global trends in greenhouse gas emissions, it is increasingly apparent that some degree of warming as well as other climate impacts will be inevitable, requiring adaptive actions at different levels and by different actor groups (Adger et al., 2005). While policy development for climate change adaptation is still at an early stage in many countries (Massey et al., 2014), it is clear that interventions for adaptation will differ fundamentally from those in mitigation. The need for deliberative policy appraisal for adaptation is even more pronounced, given the unstructured character of adaptation problems, the many uncertainties surrounding climate impacts and agents´ capacity to respond, and the often localised nature of solutions, where one-size-fits-all approaches will often fail. Not without reason, the literatures on adaptive governance and adaptive co-management emphasise notions such as participation, learning and experimentation as key means for governing socio-ecological systems in times of environmental change (Armitage et al., 2007; Armitage et al., 2008; Folke et al., 2005).

In conclusion, climate governance is undergoing a transformation in several respects. Consequently, the need for interactive appraisal methods that allow for dealing with the complexity of the issues at stake, while fostering an effective exchange of information and ideas among experts and stakeholders, is probably greater than ever. The results of this research indicate that policy exercises and policy games, while perhaps not the panacea they were described as in the early literature, can be helpful

in this regard. Yet for this field to realise its promise, the next stage of research on participatory policy appraisal needs to go beyond an uncritical application of certain methods and take a harder, more systematic and hypothesis-driven look at their respective outcomes, costs and benefits, be it in terms of learning or otherwise. Only then will this literature be able to retain its credibility also in the longer term.

Annex 1 – Policy exercise materials and evaluation instruments

Annex 1.A Sample interview guide for ex-post interviews (here for burden-sharing policy exercise with experts)

1. What did you like about the design of the policy exercise?
2. And what should have been different?
3. Do you think the time limit made the exercise more realistic?
4. At a more general level, what do you see as the key advantages of the policy exercise method? And the key drawbacks?
5. Did you find the policy exercise method a useful way of thinking about the more distant future?
6. If so, in which way was your thinking stimulated? How could it have been stimulated even more?
7. Did the ADAM policy exercise generate innovative ideas or options?
8. In how far did you feel that the policy exercise was different from other international workshops that you have attended?
9. In how far did the policy exercise format influence the way of interacting among participants?
10. In the exercise, you were asked to play a role. Is this the role you normally operate from?
11. If not, to which extent has playing the role furthered your insight in the way the actors in that role operate in reality?
12. If we were to organise another policy exercise, what advice would you give us? What should we do in the same way, what should we do differently?

Annex 1.B Guiding questions for debriefing session (here for REDD policy exercise with experts)

I. The substance of the exercise

- Outside of the heat of the moment, do the outputs of the policy exercise seem convincing? Why or why not? What conditions would make it so or not?
- Our starting point was a rather benign global climate policy scenario – a continuation of the global climate regime post-2012. However, would any of this be feasible under a more fragmented approach?
- What would you most have needed to know to make a decision or choice?
- What are the research implications/knowledge needs flowing from this exercise?

II. Experience of the policy exercise

- How was it for you? Was it a day (and effort) worthwhile spent?
- To what extent did you find the atmosphere and the interaction with participants different from a 'normal' workshop format?
- What are you taking home from this day?

III. Critique of approach and design of the exercise

- If we were to organise another policy exercise, what advice would you give us?
- Scenarios
- Teams and representation:
    - Implications of few teams;
    - Implications of no internal structure; alternatives?
- Procedures:
    - Was this most useful/important way to present the tasks?
    - What would be salient alternatives?

Annex 1.C Participant lists for expert runs of the policy exercises

Policy exercise on EU burden-sharing

Ian Bailey	University of Plymouth
Markus Brede	CSIRO
Dagnija Blumberga	Institute of Energy Systems and Environment
Javier de Cendra	Maastricht University
Bert de Vries	Netherlands Environmental Assessment Agency (PBL)
Dana Dvorakova	CEZ Group
David Ellison	Institute for World Economics
Angela Falconer	AEA Technologies
Roger Hildingsson	University of Lund, Sweden
Andy Kerr	E3 International
Kai Kühnhenn	Federal Environment Agency
Sigurd Naess-Schmidt	Copenhagen Economics
Eva Lövbrand	University of Lund, Sweden
Bert Metz	Netherlands Environmental Assessment Agency (PBL)
Till Neeff	Ecosecurities
Joop Oude Lohuis	Netherlands Environmental Assessment Agency (PBL)
Tim Rayner	University of East Anglia
Jürgen Salay	European Commission, DG Environment
Jos Sijm	Energy Research Centre of the Netherlands

Johannes Stripple	University of Lund, Sweden
Louis van Schaik	Clingendael Institute
Timme van Melle	Vu University Amsterdam

Policy exercise on REDD

Juan Pablo Castro	Climate Focus
Ernestine Meijer	Climate Focus
Maria Nijnik	Macaulay Land Use Research Institute
Benoit Morel	Carnegie Mellon University
Alexandra Morel	Columbia University
Innocent Bakam	Macaulay Land Use Research Instititute
Herry Purnomo	Centre for International Forestry Research (CIFOR)
Cordula Epple	UNEP-World Conservation Monitoring Centre
Jan Fehse	Ecosecurities/Value for Nature Consulting
Harro van Asselt	University of Oxford
Onno Kuik	VU University Amsterdam
George Dyer	Macaulay Institute for Land Use Research
Stefanie von Scheliha	Gesellschaft für technische Zusammenarbeit (GTZ)
Herman Savenije	Tropenbos International
Deepak Rughani	Biofuelwatch
Angelica Mendoza	Netherlands Environmental Assessment Agency (PBL)
Harko Koster	WWF Netherlands

Annex 1.D Scenarios, policy exercise on burden-sharing

Burden-sharing exercise, Scenario 1: "From Kyoto to Kingston"

We are in the year 2018. Scientific evidence for climate change has increased significantly; severe climate impacts are already felt, in Europe and beyond. The post-2012 Kingston Protocol has gathered a large degree of support, but compliance so far has been variable. The US, the new green climate policy leader and the BRIC countries have embarked on an ambitious mitigation path. The EU, however, is coming to the realisation that its 2009 climate package is a failure. Effort-sharing in particular has turned largely ineffective and controversial among member states. New impetus is urgently needed for EU climate policy if the Union is to leave its imprint on the yet-to-be-agreed post-2020 climate agreement…

The IPCC Fifth Assessment Report narrowed down uncertainty margins on a number of issues. On the impact side, agriculture and tourism in Southern Europe suffered due to extensive heat waves and droughts, while recurrent flooding along the Rhine and in the Tisza basin testifies to the vulnerability of Central and Eastern Europe to climate change.

The "Kingston Protocol" was finally agreed in Jamaica in 2010, after the Copenhagen summit ended in stalemate, and has meanwhile been ratified by 183 countries. The agreement endorsed a second commitment period (2013-2020) to the Kyoto Protocol, with absolute reduction targets for all industrialised countries. No binding targets were set for non-Annex-I/B countries. Instead, they subscribed to voluntary commitments to reduce emissions through Sustainable Development Policies and Measures (SD-PAMs). Market mechanisms moved centre stage in Kingston, with sectoral baselines and credits from avoided deforestation both included in the agreement.

As to mitigation progress among the world's largest emitters, the unexpected rising star at the climate policy horizon is the US, where President Obama and his successor have sparked a technology revolution, with large-scale uptake of clean coal, nuclear energy and carbon capture and storage resulting in a sharp downturn in emissions. Its Kingston target is in close reach, whereas the performance of Canada, Australia and the EU is far less glorious. The BRIC countries, however, have taken advantage of the opportunities provided by the Kingston Protocol, and have succeeded to slow down their emissions growth. As China and Brazil are quickly approaching the 'graduation threshold' endorsed in the Jamaica accord, it seems likely that they will have to commit to binding emission limitation targets under a post-2020 agreement.

The EU has a tumultuous decade behind it. The European economy has started to grow again since 2014. The piecemeal integration of the provisions of the failed Lisbon Treaty

into European law has laid the ground for governing the EU after the 2015/2017 Balkan enlargements. Yet large disparities in wealth and political priorities complicate EU-internal decision-making. Furthermore, there is a clear feeling that the EU's 2009 climate package as a whole is unlikely to deliver on its promises. If the EU is to reach its 30% Kingston target, it will be largely due to external carbon credits. One seemingly innocuous provision in the revised EU ETS directive, the admission of domestic offsets, has severely compromised the functioning of the trading scheme and has blurred the boundaries with the effort-sharing domain (as emission reductions in non-ETS sectors were as a consequence double-counted under the ETS). Furthermore, allowances auctioned under the ETS have to a large extent been bought up by big multi-national companies, hurting smaller and less profitable industries, especially in Central and Eastern Europe.

While major revisions of other EU climate policies are also pending, this time the EU Commission first wants to get it right in the effort-sharing domain, which has proven the least successful and most controversial part of the package. Very few countries, except for the Nordics, the UK, and the Baltic states, are on track to meeting their effort-sharing targets. The challenge to design national policies that effectively contribute to structural changes in the building and transport sectors is unresolved in many member states. In 2016, a group of countries, including Poland, the Czech Republic, Slovakia and Hungary had a major fall-out with the European Commission, after the latter's decision to initiate infringement procedures for failure to comply with the 2015 and 2016 interim targets following two particularly harsh winters in Central Europe. Although the European Commission ultimately backed down, this incident severely damaged the relation between Brussels and the CEE capitals.

All this hurts the EU, the former global climate policy leader, which looks set to loose this image to the new green superpower US. If the EU wants have an impact on the upcoming post-2020 negotiations, it needs to act quickly to demonstrate to the world that it recognises the implementation failures of the past and is ready to adopt robust policies that will deliver the projected emission reductions with a high degree of certainty.

Burden-sharing exercise, Scenario 2: "Coat with many colours"

We are in the year 2018. Scientific evidence for climate change has increased significantly; severe climate impacts are already felt, in Europe and beyond. After the failed ratification of the Copenhagen Protocol, the UNFCCC has transformed into a global climate policy clearing house. In parallel, a multitude of climate initiatives and partnerships led by the new green superpower, the US, have successfully spurred

mitigation efforts, also in China, Brazil and India. The EU, however, has come to the realisation that its 2008 climate package, despite the relative success of the EU ETS, has failed on the whole. Effort-sharing in particular has proved to be largely ineffective and controversial among member states. New impetus is urgently needed if the Union is to restore its tainted image as a global climate policy leader…

The IPCC Fifth Assessment Report narrowed down uncertainty margins on a number of issues. On the impact side, agriculture and tourism in Southern Europe suffered due to extensive heat waves and droughts, while recurrent flooding along the Rhine and in the Tisza basin testifies to the vulnerability of Central and Eastern Europe to climate change.

Under heavy pressure from the EU, the Cancun Protocol was concluded in 2010. Yet when the adoption of the 2012 Cornyn Act by the US Congress closed the door to US ratification, Canada, Japan, Australia, and the BASIC countries, feeling they had overcommitted in the first place, were quick to follow. After trying to resuscitate the agreement for a while, the EU accepted the transformation of the UNFCCC into a sort of clearing-house, where countries convene to take stock of ongoing efforts and to share best practices.

Outside of the UNFCCC, a myriad of climate initiatives and low-carbon technology cooperation projects have sprung up. The first were spurred by the EU, which concluded a series of bilateral partnerships with developing countries to secure carbon credit supply. Even more important are the initiatives led by the US, where President Obama and his successor sparked a technology revolution, with large-scale uptake of clean coal (including CCS) and nuclear energy resulting in a sharp downturn in emissions. Developing countries, especially China, India and Brazil, have benefited from international technology cooperation and strengthened their domestic climate policies. Their emission growth has slowed down in recent years, leaving the EU in an unexpected laggard position.

The EU has a tumultuous decade behind it. After the financial crisis at the end of the last decade, the European economy has started to grow again since 2014. The piecemeal integration of the provisions of the failed Lisbon Treaty into European law has laid the ground for governing the EU after the 2015/2017 Balkan enlargements. Yet large disparities in wealth and political priorities complicate EU-internal decision-making. There is a clear feeling that the EU's 2009 climate package as a whole is unlikely to deliver on its promises. If the EU is to reach its unilateral 20%-by-2020 target, it will be largely due to external carbon credits. The only bright spot is the EU ETS. The third and fourth phase auctioning of allowances has created a sufficiently high carbon price (partly mitigated by external credits), which has started to pay off by transforming key industrial sectors. Liquidity of the system has improved after linking to trading schemes in Australia, Japan and New Zealand and, finally, the US. In relation to other markets, border tax

adjustments have been introduced to address carbon leakage concerns and to safeguard European industries against competitiveness losses.

While major revisions of other EU climate policies are also pending, this time the EU Commission first wants to get it right in the effort-sharing domain. Very few countries, namely the Nordics, the UK, and the Baltic states, are on track to meet their effort-sharing targets. The challenge to design national policies that effectively contribute to structural changes in the building and transport sectors is unresolved in many member states. In 2016, a group of member states, including Poland, the Czech Republic, Slovakia and Hungary had a major fall-out with the European Commission, after the latter's decision to initiate infringement procedures for failure to comply with the 2015 and 2016 interim targets following two particularly harsh winters in Central Europe. Although the European Commission ultimately backed down, this incident severely harms the relation between Brussels and the CEE capitals.

All this damages the EU, the former global climate policy leader, which looks set to lose this image to the new green superpower US. If the EU wants to restore its climate policy reputation, one of the few areas where it can still engender the population's enthusiasm, it needs to act quickly to demonstrate to the world that it recognises the mistakes of the past and is ready to adopt robust policies that will deliver the projected emission reductions with a high degree of certainty.

Annex 1.E Role descriptions, policy exercise on burden-sharing

Germany in 2018

Facts & Figures (year 2016)

		Relative position in EU-33
National emission trend (ETS/non-ETS) compared to 2005	-8%	
National emissions in million tCO2	924	1st
Emission trend in non-ETS sectors, compared to 2005 (excluding use of carbon credits)	Target: -14% Actual: -6%	
featurePer capita emissions in tCO2/person	11,2	4th
GDP development over the period 2008-2018	+20% (EU average)	
GDP in 1000 million Euros/yr	3.018	1st
GDP per capita in Euros/yr	36.493	10th

Over the last 10 years, Germany has seen robust economic growth, has consolidated its leading role in Europe both economically and politically, and has continued its transformation towards a services economy.

Throughout the last decade, climate policy has remained a priority in Germany. By and large, the country is on track in the effort-sharing domain, however relying, to a large extent on the purchase of external carbon credits and on excess reductions transferred from other member states. Domestically, while policy measures on insulation, heating and cooling have proven moderately effective, attempts to encourage the modal shift in transport seem to have been failing so far. In the EU-ETS sectors, participation in the scheme has turned out costly for some parts of German industry, with the metalworking and aluminium industry suffering most. Large amounts of external credits combined with allowances bought from other member states' auctions have made compliance possible, but have angered smaller installations (and governments) in other member states, especially Poland.

At this moment, Germany holds the EU presidency and considers it of high priority to push the EU on an ambitious climate policy, without, however, excessively hurting the competitiveness of its own export-oriented industries.

Sweden in 2018

Facts & Figures (year 2016)

		Relative position in EU-33
National emission trend (ETS/non-ETS) compared to 2005	-11%	
National emissions in million tCO2	59	30th
Emission trend in non-ETS sectors, compared to 2005 (excluding use of carbon credits)	Target: -17% Actual: -12%	
Per capita emissions in tCO2/person	6,2	27th
GDP development over the period 2008-2018	+12% (EU average: 20%)	
GDP in 1000 million Euros/yr	378	9th
GDP per capita in Euros/yr	39.425	7th

Efforts to fight climate change have received high priority in Sweden over the last decade. The results are impressive – emissions have decreased sharply, renewables growth has been robust, and greenhouse gas emissions per capita are among the lowest in the EU despite a continuously high standard of living.

Especially in the effort-sharing domain, the Swedish case is one of the rare success stories in the EU. Relying on a balanced strategy of purchases of high quality carbon credits and implementing ambitious policies, the country looks set to achieve its -17% effort-sharing target. Some studies suggest, however, that climate policy efforts (both ETS/non-ETS) have taken their toll on the Swedish economy and are among the key reasons for slower-than-EU-average economic growth in Sweden. Furthermore, most of the cheaper non-ETS abatement options seem to be exhausted for the moment, with some limited potential left only in the transport sector.

The Swedish population is overall still supportive of the climate policy strategy of government, yet begins to tire of the efforts made and the costs involved. The conservative opposition party points to the lack of progress in other member states and insists that they should step up their efforts before Sweden adopts even more stringent policies.

Spain in 2018

Facts & Figures (year 2016)

		Relative position in EU-33
National emission trend (ETS/non-ETS) compared to 2005	+3%	
National emissions in million tCO2	454	10th
Emission trend in non-ETS sectors, compared to 2005 (excluding use of carbon credits)	Target: -10% Actual: +2%	
Per capita emissions in tCO2/person	9,9	12th
GDP development over the period 2008-2018	0% (EU average: 20%)	
GDP in 1000 million Euros/yr	1.101	5th
GDP per capita in Euros/yr	24.142	16th

The past decade has been rocky for Spain. The global financial crisis in 2008/2009 led to a property bust across the Iberian Peninsula, propelling the country into a deep and drawn-out recession, from which Spain has yet to recover.

On the climate policy front, the previous government at first made a serious effort to shed its laggard image after its failure to its Kyoto burden-sharing target. This translated into a renewables boom and decreasing emissions in the residential sector. Yet as the economic situation worsened and unemployment figures rose, climate policy increasingly took a backseat among other, more pressing concerns. A compliance strategy largely based on the purchase of external credits was adopted for the effort-sharing domain – which, at a time of empty state coffers, did not fare particularly well with the Spanish public. Since 2014, Spain has clearly deviated from the linear reduction path stipulated by the 2009 effort-sharing decision and is now waiting for a "letter of formal notice" from the Commission. At the same time, there are still large mitigation potentials available, as for example the much-needed substitution of the many outdated and emission-intensive vehicles in public and private transport.

Moreover, the adverse impacts of global warming on the tourism and agricultural sector have alerted the Spanish public to the threat of climate change. In next year's elections, the approach of the candidates towards climate change is likely to play a pivotal role.

Poland in 2018

Facts & Figures (year 2016)

		Relative position in EU-33
National emission trend (ETS/non-ETS) compared to 2005	+7%	
National emissions in million tCO2	425	15th
Emission trend in non-ETS sectors, compared to 2005 (excluding use of carbon credits)	Target: +14% Actual: +28%	
Per capita emissions in tCO2/person	11,4	2nd
GDP development over the period 2008-2018	+50% (EU average: +20%)	
GDP in 1000 million Euros/yr	571	7th
GDP per capita in Euros/yr	15.400	22nd

Poland has seen sustained high economic growth over the last decade, thanks to far-reaching liberalisation efforts and the introduction of a 50% tax discount for repatriates, which stimulated to large-scale technology and capital transfers. Today, the country is a dynamic, largely internationally competitive economy, with GDP per capita quickly growing towards the EU average.

As to climate policy, however, the country remains a 'problem child' within the EU. Successive governments underestimated the difficulties of complying with the requirements of the EU's 2009 climate package, especially at a time of strong economic growth and increasing purchasing power. Half-hearted attempts at policy implementation and enforcement have not had much impact on emissions, especially in the non-ETS sectors. This does mean however, that affordable abatement options, such as improving insulation, are still available.

In 2013/2014, Poland slipped into non-compliance with the effort-sharing decision for the first time. The cold winters in 2015/2016 increased the compliance gap, prompting the Commission to initiate infringement proceedings, which are currently pending. At this moment, the Polish government is not showing much effort to meet the 2020 effort-sharing target. Taking into account the high level of economic growth and the extraordinary weather, it takes the stand that in hindsight, the targets were unreasonably strict.

Overall, however, your government and the Polish public remain supportive of European integration. Especially given the strained relations with Russia, Poland needs EU backing

and thus does not want to be perceived as a complete spoiler in an area where high profile EU action is important.

European Commission in 2018

You are senior officials of the Directorate General for Climate Change (DG CC) in the European Commission. DG CC was established in 2012 in an attempt to improve the cross-sectoral coordination and mainstreaming of climate policies.

As Commission officials, you see your role as that of honest brokers, who strive to maintain the reputation of the EU as a frontrunner in global climate policy (one of the few fields where it still manages to inspire), while keeping in mind the EU's fundamental objective of enhancing social and economic cohesion across Europe.

The adoption of the 2009 climate package was an important success for the Commission, yet one that came at a price; CEE countries, but also others, complained about the lack of transparency of the proposal and the lack of time to fully comprehend the issues at stake. Trust was further shattered when the Commission rather unexpectedly tightened its definition of leakage-prone industries, effectively limiting the extent of free allocation, for the EU-ETS fourth trading period. Finally, voices, including within the Commission, doubt whether it was a wise decision to initiate infringement procedures against the group of CEE "effort-sharing offenders" in 2016, and suggest that a more cooperative approach might have been more productive.

With the failure of the EU to achieve its 2020 goals increasingly obvious, it is of crucial importance now to send a strong signal both to the European public and the international community on the EU's commitment to fight climate change. A key problem in implementing the 2009 effort-sharing decision was the unreliable and irregular reporting by member states, which made assessments of progress towards target extremely difficult. This certainly wasn't helped by the large degree of flexibility (both internal and external) that member states were given in meeting their commitments. It is up to you now to make a first proposal that stands a chance of being accepted by member states yet provides a more robust basis to ensure that ambitious targets for the non-ETS sectors are actually being met.

Annex 1.F Policy element cards, policy exercise on burden-sharing

Policy elements	Option 1	Option 2	Option 3	Option 4	Option 5
Base year	1990	2005	2016	Average yearly emissions between _____ and _____	
Criteria for burden-sharing	Converging per capita emissions in effort-sharing sectors by 2030	Relative GDP per capita of member states	Equal percentage emission reductions in effort-sharing sectors in all member states	Least cost reduction for the whole EU in non-ETS sectors	
Compliance and enforcement	In case of non-compliance, the infringement procedure under European law applies.	In case of non-compliance, a member state has to pay a penalty equivalent to the fine under the EU ETS. Penalties shall be used for _____	In case of non-compliance, an amount equivalent to the excess reductions is deducted from the subsequent auction of allowances of that member state under the EU ETS.	Authority administering the non-compliance procedure: _European Commission _European Environment Agency _newly established European Climate Agency	
Internal flexibility	"Banking" - A member state can carry over excess emission reductions up to _____% of its overall GHG limit to the subsequent year.	"Borrowing" - A member state can carry forward emission reduction obligations up to ____% of its overall GHG limit to the subsequent year.	Member states shall comply with their targets in a linear manner.	Member states shall comply with the interim target set for the year ____ (and ____).	A member state can trade excess reductions in the effort-sharing sector with other member states.

Interaction with EU ETS	There is no exchange possible between emission reductions/allowances under the EU-ETS and under effort-sharing.	A member state can decide to allocate excess emission reductions achieved in the effort-sharing sectors for auctioning under the emissions trading system	A member state can decide to reduce its amount of allowances for auctioning under the EU ETS, and use them for compliance under the effort-sharing decision instead.	Coverage of the EU ETS will be extended to….. [ ] Transport [ ] Land use [ ] _____	
Quality of external credits	All credits admissible under the UNFCCC can be used for compliance under the effort-sharing decision..	No credits from [ ] nuclear power [ ] large hydro [ ] CCS [ ] HFC-23 [ ] _____ are admissible for compliance under the effort-sharing decision. compliance are discounted by ____%	Non-project-based carbon credits (e.g. from avoided deforestation or sectoral approaches) that are used for	At least ____ % of the credits used for compliance must be 'Gold Standard'. This percentage grows annually by ____ %.	Only credits that were awarded the Gold Standard or an equivalent quality standard can be used for compliance.

Limits to external credits	External carbon credits are not admissible for compliance under the effort-sharing decision.	Every member state has a yearly quota of ____ % of external carbon credits at its disposal. It expires if not used in that same year	Every member state has a yearly quota of _____% of external carbon credits at its disposal. Unused quota can be banked until the end of the compliance period.	Every member state has a yearly quota of ____ % of external carbon credits at its disposal. If it does not use its quota, the remaining quota may be passed on to other member states.	
Optional: new policy initiatives	Transport:	Agriculture	Building/residential sector:	Waste management	Small- and medium-sized enterprises

Annex 1.G Scenario, policy exercise on REDD

Kyoto, Cape Town and beyond

We are in the year 2015. COP-21 has just opened in Santiago de Chile. After a pause for breath after the arduous negotiations at COP-17 in Cape Town in 2011, where a deal had finally been reached on a post-2012 climate agreement, the eyes of the world are once again on the climate negotiations, and on REDD+ in particular.

What had actually been decided at Cape Town? In a nutshell, a set of decisions addressing all four elements of the Bali Action Plan endorsed what came to be known as the 'Copenhagen approach' – bottom-up emission reduction and stabilisation pledges by developed and developing countries, which were annexed to the agreement and detailed in individual country submissions. In total, developed country pledges amounted to an aggregate 30% cut by 2020 compared to 2000 levels. Quantification of the – in parts rather ambitious – actions pledged by developing countries is less straightforward, but on the whole, it is doubtful whether the commitments made in Cape Town will be sufficient for keeping the global warming below 2 degrees Celsius until the end of this century. Cape Town further saw the establishment of a Global Green Climate Fund (GGCF) under the UNFCCC, with separate funding windows for adaptation, mitigation, and REDD+. The GGCF is to be capitalised from 'appropriate, new and additional contributions' by developed countries, with the aim of approaching a funding volume of 60 billion USD annually by 2020. At this moment in 2015, the latter target is still a long way off. Total pledges to the GGCF currently amount to almost 20 billion, and the disbursement details of the funding windows are still to be worked out in further COP decisions.

With regard to REDD+, parties in Cape Town agreed that its scope should encompass deforestation, forest degradation and the enhancement of existing carbon stocks and that the scale of implementation should mainly be national. The REDD+ decision in Cape Town further endorsed a historical approach to baseline-setting, with the possibility of adjustments where appropriate. Since agreement on other aspects proved elusive at this point, Parties decided on an extended pilot and trial period for REDD+, with a view to improving data availability and gaining further experience with project implementation. COP-21 was called upon to review progress and consider the design of the funding window for REDD+ under the GGCF, negotiations on which had broken down previously over the sources and necessary scale of funding as well as disbursement rules. Observers anticipate that discussions on the REDD+-window of the GGCF will also have a pioneering character for the other funding windows, where progress has been even slower.

While the discussions under the UNFCCC were inching forward over the past few years, REDD+ developments outside of it have kept up at a steady pace. In total, about 12 billion USD has been spent on readiness efforts through multilateral and bilateral channels. The REDD+ partnership in particular has proven its value as an effective platform for information exchange and donor coordination. However, there has also been evidence of duplication of efforts and of squandered funds. Scandals in Malaysia and the Democratic Republic of Congo, where funding was received for the same REDD-readiness activities from several initiatives, have increased political momentum for a concerted global approach to REDD+.

At the same time, costs for monitoring technology have come down rapidly over the last years and data availability has improved across many tropical countries, making the widespread implementation of REDD+ activities less of a distant dream. Progress with institution-building, however, has been more uneven. While some Latin American countries have their national REDD+ administrations up and running and their REDD+ strategies and monitoring plans in place, in others, there is still a lot of work ahead.

Looking at global trends in tropical deforestation, the last five years have brought about some progress, but a real reversal in global deforestation trends has not yet been achieved. Worldwide economic recovery after the crisis at the end of the last decade has increased global demand for timber and spurred deforestation in many countries. Success stories, for example in some regions of Brazil, are largely offset by accelerating deforestation in the Congo Basin, where increased infrastructure investments combined with a lack of political accountability have attracted Asian logging companies at a large scale.

The international community thus needs progress on international REDD+ policy. Given that prices in the world's regulatory carbon markets are currently low, there is little appetite for an immediate market linkage of a global REDD+ crediting scheme. Even though more and more countries (including the United States) are making progress towards implementing domestic emissions trading schemes, it seems unlikely that there will be sufficient demand in the short term to accommodate a large amount of REDD credits.

Therefore, the task for the Santiago summit is to establish the basic modalities for a global REDD+ fund. The Parties moreover still need to decide whether this is an interim or a long-term option for REDD+ financing, and, in case a phased approach is chosen, how to engineer the transition to the integration of REDD into the carbon market. On this day, Brazil, who is particularly keen on progress on a REDD+ fund, has invited some countries for an informal meeting at the margins of negotiations. Its goal is to develop a proposal for a REDD+ fund that might be able to command broad-

based support among Parties and thus guide the further negotiations on REDD+ in Santiago de Chile.

Annex 1.H Role descriptions, policy exercise on REDD

United States in 2015

The United States in 2015 have finally begun to come to terms with their new role in a multi-polar world. After a phase of deep disorientation and anger at the lost causes of Afghanistan and Iraq as well as the economic crisis, the general mood in the US has reverted back to American 'can do' optimism. Overall, the new US paradigm emphasises moral and technological leadership over military and economic dominance.

After the first years of the new decade, which will be remembered as the 'dark ages of partisan politics' in Washington DC, things started to turn for the better after President Obama was reelected in 2012. A vitriolic Republican electoral campaign had pushed large parts of the independent voters back to the Democrats. The new, more bipartisan-minded Congress adopted long-needed legislative reform on entitlements that laid the basis for a healthy federal budget over the coming decades. The US economy has picked up as well, although growth rates compare poorly to those in the emerging economies and unemployment figures never returned to pre-crisis levels.

Meanwhile, Barack Obama, nearing the end of his second term and concerned about his legacy, has turned his attention back to global issues, chief among them climate change. Several severe natural disasters plausibly linked to climate change and the emergence of a powerful broad-based pro-climate movement of green industry, insurance companies and NGOs have convinced a majority of the public that action is needed and in the US national interest. Obama's domestic climate policy record had been exceedingly meager in his first term. Now, in 2015, the passage of a bill on a federal emissions trading scheme finally seems within reach (although only covering utilities). The bill before Congress is likely to incorporate substantial provisions on REDD+ activities and offsets. Hopes are high that this bill, in combination with a policy package on cleaner transportation, will prompt the much-needed turnaround in US emission trends that has failed to materialise thus far.

Despite this hopeful outlook, the international community – and particularly developing countries – are frustrated by the lack of results of US climate policy hitherto. Not only has the US so far failed to engage in substantial efforts to reduce

emissions at home, its record on international climate financing is equally meager. Whereas the US continues to be involved in various bilateral clean tech and REDD+ partnerships, the disbursements of US funds to the Global Green Climate Fund (GGCF) has been consistently behind schedule and below promised amounts.

Obama wants to make the Santiago summit a success and considers REDD a pivotal piece of the puzzle. However to prevent domestic backlash at third-world corruption and UN inefficiency, the US administration would like to see any fund implemented by a trusted international institution, strictly performance-based and with stringent MRV rules. Moreover, Congress is unwilling to commit long-term funding from taxpayer money. Rather, it wants carbon markets and the private sector to generate the necessary funds in the medium term.

EU in 2015

After protracted economic stagnation for several years that nearly brought down the Euro zone, the EU is finally enjoying some economic growth again. Most member states' financial stability has significantly improved, even though the degree to which structural reforms have been implemented varies largely across the Community. Politically, the EU has entered a consolidation phase, with no major institutional reforms or big political projects in sight. Further increase in membership has equally been put on ice for the moment, except for the upcoming Balkan enlargement.

The EU's economic and political stagnation, along with the continued rise of big non-Western powers has undoubtedly dented its standing in the world. Fighting climate change however provides a major narrative in Brussels' attempt to connect to Europe's peoples. Learning from the debacle in Copenhagen, the EU has taken steps to reform its international climate diplomacy. In institutional terms, climate change Commissioner Lord Stern now leads negotiations for the EU delegation in much the same way as his counterpart does in the trade arena. In his attempt to shape international negotiations, he has relied on sticks as well as carrots, effectively supported by the European External Action Service. The EU has worked very productively with the more progressive developing countries towards brokering the Cape Town accords. It is not quite clear to what extent the EU's increase of its emissions reduction target to 30% played a role in securing agreement in Cape Town.

The EU has also continued to make progress in cutting its emissions at home. The extent of the reliance of some countries on external carbon credits for compliance has however drawn substantial criticism. On the other hand, it has been European demand that has prevented prices for credits to fall even further. A review of the EU Emissions Trading Scheme (EU ETS) is currently ongoing. It will among other things

address options for the inclusion of REDD credits in the scheme after 2020, though nothing has been decided yet. The weakest part in the EU's international climate strategy is financing. The EU Commission continues to lack the necessary funds on its own to sway developing countries' positions. At the same time, an EU-internal burden-sharing formula for climate financing that the Commission had advocated never came to pass, effectively leaving the brunt of the burden on the most progressive EU Member states. Some member states have earmarked specific shares of their auctioning revenue for REDD+, but overall financing sources remain volatile and insufficient.

REDD+ continues to be a major negotiation priority for the EU and is considered an issue on which progress can be made. EU member states have been active in all major REDD+ initiatives and have engaged in a host of bilateral partnerships, but increasingly see the need for streamlining international efforts, also with a view to the prospect of including REDD credits into the EU ETS. The Community stresses that its priority is to design a 'robust' and economically efficient mechanism with stringent monitoring requirements, and emphasises the importance of environmental and social safeguards. While the EU remains somewhat cautious about a market-based approach to REDD+, most member states acknowledge that this will be the only way for leveraging the funds necessary in the longer term.

India in 2015

India has done particularly well out of the global economic recovery. A young and growing workforce is boosting innovation across a range of sectors and is driving domestic consumption, resulting in good economic prospects also in the medium term. A self-confident middle class has grown in numbers and spending power. Economic progress however has failed to trickle down to large parts of rural India. The challenges of alleviating widespread poverty remain enormous, particularly in the face of growing environmental pressures, such as water and land scarcity. At the same time, the complex institutional setting in the world's biggest democracy makes political reform difficult to achieve.

In the international arena, institutional reform in the Bretton Woods system and the United Nations has increased India's voice in the world. The country has moreover freed itself from the grip of its strategic alliance with China in the realm of global climate policy. With the latter increasingly seen among developing countries as defending its own interests rather than those of the South at large, India has managed to establish itself as a G77 leader of sorts in the climate negotiations. Even though the Indian establishment is frustrated with the limited progress of

industrialised countries in cutting emissions as well as in delivering climate finance, it sees little alternatives to a multilateral solution to the climate crisis.

In its national low-carbon growth plan, India has focused on moving away from its coal-reliant energy base. Severe climate impacts felt in India – massive agricultural losses as a consequence of several erratic monsoon reasons and repeated large scale flooding in the South – have moved the Indian government towards a more proactive stance on domestic climate policy. The country is internationally recognised as one of the most successful examples for decoupling economic growth from carbon intensity. Moreover, it has recently made significant strides in increasing re-forestation. Its limited forest cover however implies that India nationally stands to gain little from a REDD+ approach focusing on reducing deforestation and forest degradation. Since India continues to be a strong player in the CDM and is well-positioned to command a significant share of the emerging market for sectoral mechanisms, it is keen on preventing REDD from crowding out other international offsets. At the same time, India recognises the important pilot role that the operationalisation of the REDD window may have for other parts of the Global Green Climate Fund (GGCF).

India's primary objective in this negotiation is therefore to set a precedent for a balanced governance structure for climate financing, un-bureaucratic access and legally binding funding sources. India hopes that the Santiago summit will mark an end to the conventional aid paradigm that has dominated discussions on climate finance for too long, and that it will result in a new financial architecture on climate change. This architecture should respect all partners' sovereign equality, and it should be based on the principle of common but differentiated responsibility. The GGCF should thus combine national ownership of funds with non-intrusive international verification mechanisms, and its setup should reflect historic responsibility for climate change.

Brazil in 2015

The last half a decade has been a good one for Brazil. Economically, it seems that South America's largest economy has finally started to live up to its eternal label as the 'country of the future'. Having weathered the global financial and economic crisis in 2008/2009 rather well, it has since started to grow again with an annual rate averaging almost five per cent over the period 2010-2015. President Rousseff's continuation of the poverty alleviation agenda of her predecessor Lula has ensured that the benefits of economic growth reach a larger part of the population than previously; this is at the root of her reelection in 2014. Finally, Brazil's hosting of two big sport events, the 2014 FIFA World Cup and the 2016 Olympics, is boosting the country's self-confidence and national pride.

Good news for Brazil's environmentalists as well as for the global climate has been Rousseff's stern stance on fighting deforestation in the Amazon. Her efforts in this regard have largely paid off: although more than half a million hectares of forest still disappear every year, the deforestation rate has dropped drastically compared to the beginning of the 2000s, the president's headline objective of halting net deforestation by 2020 seems possible. This, especially compared to other tropical forest countries, relatively bright picture is due to progress at many levels: strong forest monitoring systems have improved further, and even small-scale deforestation activities have now become detectable by INPE, the national space agency. Good progress has been made in cleaning up the land registry and in clarifying tenure rights, leading to more sustainable forest management on privately managed lands. The enforcement of the forest code by the rather corrupt judiciary in the Amazon region remains the problem child, however, which threatens the success of the policies in the longer term.

Yet overall, given its record in cutting emissions from deforestation, Brazil belongs to the small group of high performers in the global climate regime, and appears on track towards meeting its 2020 national emission reduction goal. The country continues to be a heavyweight in the CDM market (which, however, given the modest caps of industrialised countries is less dynamic than in the pre-2012 period). Its industries are gearing up to get involved in the sectoral crediting mechanisms, the modalities of which are currently still under development. As for REDD+, the country has left the REDD-readiness phase behind. Its national REDD+ administration is up and running, and its monitoring systems and REDD+ strategy are in place – even though the latter met with controversy since various stakeholder groups felt that they had not been adequately consulted in the process.

Overall, Brazil is proud of its pioneering role in REDD+. Although it has benefited from the various bilateral REDD+ partnerships it has concluded over the last years, it stresses the need for a different scale of stable, predictable, long-term funding for REDD+, which, in its view, only a global, strictly performance-based approach can deliver. Opposition to the large scale use of REDD+ credits by industrialised countries for compliance purposes continues to be a cornerstone of the Brazilian position on REDD+, especially in the face of slumping demand on the world's carbon markets. A gradual inclusion of REDD into the carbon market, however, is not out of the question for a number of high-ranking Brazilian government officials. Brazil is very keen for progress to be made on REDD+ in Santiago de Chile, and has therefore invited some Parties to today's informal meeting, which it hopes will facilitate a constructive exchange and lead to a joint vision on a global REDD+ fund that can guide the negotiations on REDD+ in Santiago.

Cameroon in 2015

Cameroon has seen solid economic growth over the past few years, benefiting from more stable prices for oil and commodity exports since the recovery of the world economy. Asian emerging economies have made big infrastructure investments in the country, resulting in the construction of a deep sea water port and a host of power and mining projects, among other things. All these developments, however, have primarily benefitted the national elites. President Biya continues to soldier on in his 33rd year in office, largely oblivious to the grievances of his people. The country suffers from high unemployment and can offer only very limited prospects to its large young population.

Meanwhile, infrastructural expansion and large-scale logging by domestic and foreign companies have not been without consequences for Cameroon´s vast and biodiversity-rich forests. Deforestation is also beginning to have a significant environmental impact. In the north of the country in particular, deforestation has been blamed for increasing soil erosion, desertification, and reduced quality of pastureland. While some REDD+ demonstration activities implemented by international NGOs have met with moderate success, the country as a whole still lacks an effective forest conservation program. Cameroon still has some way to go with regard to REDD-readiness. On the bright side, bilateral and multilateral donors have helped to improve forest monitoring, and have strengthened long-needed robust information on the country's forest reserves, including deforestation rates and drivers. However, institution-building for REDD+ has progressed only in fits and starts, hampered by bureaucratic infighting and a lack of donor coordination.

Turning to the international arena, Cameroon is very disappointed with the progress of the global climate change regime and of REDD+ in particular. Industrialised countries have failed to make credible efforts to reduce their emissions at the scale needed for averting the worst impacts of climate change on the South. At the same time, REDD+ has not yet delivered benefits on the hoped-for scale. The country's key goal for the Santiago summit is thus to ensure predictable, long-term financing of the Global Green Climate Fund (GGCF) coupled with direct access to all funds raised. For this purpose, Cameroon wants industrialised countries to replace meagre voluntary payments with substantial contributions raised on the basis of legally binding agreements. At the same time, it seeks to avoid overly bureaucratic procedures for fund disbursement and heavy-handed involvement by the international community in matters of national sovereignty. Rather than hiring expensive foreign consultants, money should be invested into domestic institutions. Moreover, Cameroon would like to see compensation not only for reduced deforestation and forest degradation, but also for the enhancement of carbon stocks at the earliest possible moment.

Annex 1.I Policy element cards, policy exercise on REDD

Policy elements	Option 1	Option 2	Option 3	Option 4	Option 5
Capitalisation and fund replenishment	Voluntary contributions	Defined bugetary contributions by industrialised countries, based on _	Share of proceeds on flexible mechanisms (similarly to the adaptation levy under the CDM)	Share of revenue from auctioning [national] [international] emission allowances	Levy on international aviation and maritime transport
Accounting for co-benefits	All [or _% of] REDD units must meet specified co-benefits minimum standards to be eligible for payment.	All [or _% of] REDD units must be certified for high co-benefits by an independent third-party verifier to be eligible for payment.	Relative co-benefit performance (measured in environmental and social 'value points') is factored into REDD unit price, resulting in multiplying the unit price for better-than-average performance.	<Auctioning approach> Separation into different auctioning tranches, moving from higher to lower co-benefits REDD units [as evidenced by certification through a specific standard] [as evidenced by the achieved number of value points]	

Eligibility criteria/How to prioritise specific funding needs in view of constrained overall funding	[REDD units from] Countries with _low income _good governance _high deforestation pressure _good past performance in deforestation _proven ability for co-financing are prioritised.	*<Auctioning approach>* Auctioning in dedicated thematic/ special needs/geographic tranches.	*<Windows-based approach>* Dedicated funding windows for _LDCs _HFLDs _....	
How to deal with the potential need for upfront financing?	Full upfront payment for first financing period, for future periods as a function of performance during previous periods.	No upfront payments.	Upfront payments [up to _% of the estimated total] are issued as loans, and subsequently transformed into grants if targets are met.	Upfront payments are made, discounted by _% for delivery risk.
Verification	Standing international panel	Expert review teams	Unilateral MRV by REDD-country	Accredited private entities (similarly to Designated Operational Entities under the CDM)
Fund administration	Global Environment Facility	World Bank	High-Level Body, composed of _	UN-REDD
Frequency of	Continuous	Annual	Biannual	Five-yearly

disbursement rounds					
Determination of payment size for REDD+ units/ implementation of REDD+ policies	<Windows-based approach> Based on countries' own estimate for implementation costs [subject to review by _]	<Auctioning approach> REDD fund purchases MRV-based REDD units from the lowest bid by a REDD country , then the second-lowest, etc, until the tranche is exhausted.	<Windows-based approach> REDD unit price is based on carbon market price [_year average] [discounted by_]	[International panel] [High level body] […] determines size of payment, based on the implementation strategy submitted by the REDD+ country.	
[Optional] Recommendation on transitioning from fund approach to integration of REDD into the global carbon markets	No transition envisaged. REDD+ fund is the long term vision.	REDD+ fund ceases operation by _	REDD+ fund acts as 'investment reserve', purchasing yet-to-be-created REDD units upfront, then banks MRV-based credits until the governing body decides that credits can be sold to the carbon market [subject to minimum price] [up to a maximum amount]	Dual-markets approach as next step (CCAP, 2007) post REDD+ fund. Creation of a new, separate REDD carbon market, where Annex I parties specify at the outset how many , and from which countries REDD units they will purchase as offsets, subject to a maximum amount determined by the COP.	TDERM approach (Hare et al., 2008) as next step post REDD+ fund. Industrialised ocuntries provide financing for REDD units proportional to their overall emission allowances in the commitment period.

Annex 2 – Learning measurements: supporting materials and results

Annex 2.A Coding systems for concept map analysis

Coding system, policy exercise on burden-sharing

Label	Explanation	Sample responses on participants' maps
Actors	Any type of actor, from governments to the private sector	Member states; industry
Physical climate	Physical aspects of climate change	Temperature increase; catastrophe; greenhouse gases
Criteria	Criteria for burden-sharing	Marginal costs of emission reduction; GDP/capita
Economic aspects	Economic factors affecting EU climate policy/burden-sharing	Oil supply; prices
EU ETS	Aspects relating to the EU ETS or the interrelation between the EU ETS	Can sectors be integrated into EU ETS; external credits; trading
Policy design & implementation	Options for policy design and implementation of EU burden-sharing	Compliance regime; monitoring; penalties
International context	International context of EU burden-sharing	Post-Kyoto regime; links to other markets
Politics & negotiations	Aspects relating to rules and dynamics of decision-making as well as negotiation skills, strategies, constraints	Deals between member states;
Policy objectives	General objectives in policy-making	Transparency; flexibility; meet EU goal
Temporal aspects	Aspects relating to the temporal dimension of EU burden-sharing	Path to final year; reducing emissions later on
Meta-reflections	Reflections on the design of the policy exercise [only	Scenario realism; number of countries

Coding system, policy exercise on REDD

Label	Explanation	Sample responses on participants' maps
Actors	Any type of actor, from governments to the private sector	Governments; donors; indigenous people
Allocation/disbursement	Aspects relating to the distribution of funds from REDD	Adequate access for all parties; equity; performance-based
Financing and sources	Funding streams and options for replenishing the global REDD fund	Capitalisation modalities; mandatory contributions; whose money
Co-benefits of REDD	Additional benefits from REDD beyond GHG reductions	Biodiversity; poverty alleviation; amounts based on carbon only or premium for co-benefits?
Environmental performance of REDD	Rule-making or implementation aspects related to safeguarding or increasing the environmental integrity of a REDD mechanism	Performance criteria and monitoring; stock inventories; verification needs
Governance of the mechanism	Aspects relating to the steering and management of a REDD mechanism	Representation in decision-making; voting modalities
International context	International policy context to and external drivers shaping the development of REDD	Climate agreement? Future forest convention; LULUCF
REDD implementation	Issues relating to the execution and management of REDD activities on the ground	Land tenure; lacking capacity in local institutions
Politics & negotiations	Aspects relating to rules and dynamics of decision-making as well as negotiation skills, strategies, constraints	Power games; reveal true policy preferences
Policy objectives	Objectives in policy-making pertaining to REDD	Short-term: work on capacity in LDCs; define key REDD priorities; address deforestation
Temporal aspects	Aspects relating to the (inter)temporal dimension of REDD and future prospects of the mechanism	Transition to carbon markets; financing short vs. long-term

Annex 2.B Results from concept map analysis, based on Morine-Dershimer (1993)

Table 23. Centrality and specificity scores on concept maps, experts burden-sharing exercise. [1]

	Centrality				*Specificity*		
	Mean score map 1 (X_1)	Mean score map 2 (X_2)	Change (X_1-X_2)		Mean score map 1 (X_1)	Mean score map 2 (X_2)	Change (X_1-X_2)
Actors	3.20	2.00	1.20	Actors	0.0758	0.0823	0.0065
Physical climate	4.40	3.80	0.60	Physical climate	0.0278	0.0250	0.0028
Criteria	1.80	1.60	0.20	Criteria	0.3232	0.2103	0.1129
Economic aspects	4.40	3.60	0.80	Economic aspects	0.1000	0.0618	0.0382
EU ETS	4.20	2.40	1.80	EU ETS	0.0273	0.1599	-0.1326
Policy design & implementation	2.00	1.00	1.00	Policy design & implementation	0.0563	0.1940	-0.1376
International context	2.00	1.20	0.80	International context	0.1379	0.1166	0.0213
Politics & negotiations	3.60	4.60	-1.00	Politics & negotiations	0.0361	0.0056	0.0306
Policy objectives	2.00	1.80	0.20	Policy objectives	0.1475	0.1024	0.0451
Temporal aspects	3.40	2.80	0.60	Temporal aspects	0.0682	0.0423	0.0259

[1] Since no items on the expert maps referred to the category 'meta reflections', the latter is not included into this table.

Table 24. Summary of analysis of variances. Differences of centrality and specificity scores between pre- and post-test, experts burden-sharing exercise. [1]

	Centrality			*Specificity*	
	df	F		df	F
Actors	1,8	1.26	Actors	1,8	0.02
Physical climate	1.8	0.18	Physical climate	1.8	0.01
Criteria	1.8	0.04	Criteria	1.8	0.82
Economic aspects	1.8	0.40	Economic aspects	1.8	0.12
EU ETS	1.8	2.49	EU ETS	1.8	6.14**
Policy design & implementation	1.8	3.33	Policy design & implementation	1.8	4.03*
International context	1.8	1.88	International context	1.8	0.09
Politics & negotiations	1.8	0.52	Politics & negotiations	1.8	1.46
Policy objectives	1.8	0.03	Policy objectives	1.8	0.48
Temporal aspects	1.8	0.28	Temporal aspects	1.8	0.44

[1] Since no items on the expert maps referred to the category 'meta reflections', the latter is not included into this table.

* p < .10, ** p < .05, *** p < .01

Table 25. Centrality and specificity scores on concept maps, students burden-sharing exercise.

	Centrality				*Specificity*		
	Mean score map 1 (X_1)	Mean score map 2 (X_2)	Change (X_1-X_2)		Mean score map 1 (X_1)	Mean score map 2 (X_2)	Change (X_1-X_2)
Actors	2.33	2.80	-0.47	Actors	0.2414	0.0859	-0.1555
Physical climate	3.67	3.93	-0.27	Physical climate	0.0467	0.0183	-0.0284
Criteria	2.45	2.05	0.40	Criteria	0.1481	0.1439	-0.0042
Economic aspects	3.42	3.68	-0.27	Economic aspects	0.0517	0.0285	-0.0232
EU ETS	3.70	2.32	1.38	EU ETS	0.0320	0.1448	0.1127
Policy design & implementation	2.60	2.25	0.35	Policy design & implementation	0.1297	0.1726	0.0429
International context	3.58	3.60	-0.02	International context	0.0389	0.0329	-0.0060
Politics & negotiations	2.60	2.07	0.53	Politics & negotiations	0.1365	0.2302	0.0936
Policy objectives	2.15	2.47	-0.32	Policy objectives	0.1404	0.0795	-0.0609
Temporal aspects	3.55	3.15	0.40	Temporal aspects	0.0336	0.0496	0.0160
Meta-reflections	4.27	4.42	-0.15	Meta-reflections	0.0009	0.0138	0.0129

Table 26. Summary of analysis of variances. Differences of centrality and specificity scores between pre- and post-test, students burden-sharing exercise.

	Centrality			*Specificity*	
	df	F		df	F
Actors	1,118	2.18	Actors[1]	1, 83.57	21.90***
Physical climate	1,118	1.13	Physical climate	1,118	3.04*
Criteria	1,118	2.05	Criteria	1,118	0.02
Economic aspects	1,118	0.89	Economic aspects	1,118	1.95
EU ETS[1]	1, 111.50	29.00***	EU ETS[1]	1,82.51	20.98***
Policy design & implementation	1,118	1.50	Policy design & implementation	1,118	2.21

International context	1,118	0.00	International context	1,118	0.18
Politics & negotiations[1]	1,118	3.21*	Politics & negotiations[1]	1,104.95	5.89**
Policy objectives	1,118	1.19	Policy objectives[1]	1,101.38	8.53**
Temporal aspects	1,118	2.09	Temporal aspects	1,118	1.59
Meta-reflections	1,118	1.04	Meta-reflections	1,118	1.75

[1] Welch-test (robust against violation of equal variances) reported as the assumption of homogeneity of variance was violated for ANOVA.
* p < .10, ** p < .05, *** p < .01

Table 27. Summary of analysis of variances. Between-group (expert/student) differences in centrality and specificity scores at pre-test, burden-sharing exercise.

	Centrality			*Specificity*	
	df	F		df	F
Actors	1,63	1.44	Actors[1]	1,10.31	11.67
Physical climate	1,63	1.28	Physical climate	1,63	.126
Criteria	1,63	.77	Criteria	1,63	4.45**
Economic aspects	1,63	1.85	Economic aspect[1]	1,4.17	.66
EU ETS	1,63	.70	EU ETS[1]	1,5.19	.88
Policy design & implementation[1]	1,5.12	.35	Policy design & implementation[1]	1,12.65	6.20**
International context	1,63	5.67**	International context	1,63	6.06**
Politics & negotiations	1,63	1.53	Politics & negotiations[1]	1,12.30	12.30**
Policy objectives	1,63	.04	Policy objectives	1,63	.013
Temporal aspects	1,63	.06	Temporal aspects	1,63	1.31

[1] Welch-test (robust against violation of equal variances) reported as the assumption of homogeneity of variance was violated for ANOVA.
* p < .10, ** p < .05, *** p < .01

Table 28. Summary of analysis of variances. Between-group (expert/student) differences in centrality and specificity scores at post-test, burden-sharing exercise.

	Centrality			*Specificity*	
	df	F		df	F
Actors	1,63	.83	Actors	1,63	.01
Physical climate	1,63	.04	Physical climate	1,63	.10
Criteria	1,63	.43	Criteria	1,63	.99
Economic aspects	1,63	.01	Economic aspects	1,63	1.17

EU ETS[1]	1,4.61	.92	EU ETS	1,63	.04
Policy design & implementation[2]	1,63	3.04*	Policy design & implementation[1]	1,63	.08
International context[1]	1,13.76	75.52***	International context	1,63	6.20**
Politics & negotiations	1,4.53	9.14**	Politics & negotiations[1]	1,48.63	61.81***
Policy objectives	1,63	.80	Policy objectives	1,63	.32
Temporal aspects	1,63	.20	Temporal aspects	1,63	.05

[1] Welch-test (robust against violation of equal variances) reported as the assumption of homogeneity of variance was violated for ANOVA.
[2]Homogeneity of variance violated for ANOVA, but Welch-test could not be conducted as the variance in one group was 0.
* p < .10, ** p < .05, *** p < .01

Table 29. Centrality and specificity scores on concept maps, experts REDD exercise.

	Centrality				*Specificity*		
	Mean score map 1 (X_1)	Mean score map 2 (X_2)	Change (X_1-X_2)		Mean score map 1 (X_1)	Mean score map 2 (X_2)	Change (X_1-X_2)
Financing and sources	1.78	1.56	0.22	Financing and sources	0.1356	0.1944	-0.0589
Allocation/ disbursement	1.33	1.44	-0.11	Allocation/dis bursement	0.2233	0.2178	0.0056
Governance of mechanism	2.67	2.11	0.56	Governance of mechanism	0.0622	0.1278	-0.0656
Env. performance of REDD	3.00	3.00	0.00	Env.performa nce of REDD	0.0867	0.0844	0.0022
Policy objectives	3.11	2.67	0.44	Policy objectives	0.0467	0.1200	-0.0733
Actors	2.89	4.22	-1.33	Actors	0.0933	0.0089	0.0844
Implementa-tion of REDD	2.67	2.78	-0.11	Implementa-tion of REDD	0.1144	0.0811	0.0333
International context	2.44	2.89	-0.44	International context	0.1689	0.0711	0.0978
Co-benefits of REDD	4.11	3.78	0.33	Co-benefits of REDD	0.0122	0.0167	-0.0044
Temporal aspects	3.11	2.78	0.33	Temporal aspects	0.0567	0.0822	-0.0256
Politics & negotiations	4.33	4.44	-0.11	Politics & negotiations	0.0000	0.0000	0.0000

Table 30. Summary of analysis of variances. Differences of centrality and specificity scores between pre- and post-test, experts REDD exercise.

	Centrality			*Specificity*	
	df	F		df	F
Financing and sources	1,16	.09	Financing and sources	1,16	1.45
Allocation/disbursement	1,16	.09	Allocation/disbursement	1,16	.007
Governance of mechanism	1,16	.44	Governance of mechanism	1,16	1.96
Environmental performance of REDD	1,16	.00	Environmental performance of REDD	1,16	.003
Policy objectives	1,16	.32	Policy objectives	1,16	1.26
Actors	1,16	4.84**	Actors[1]	1,8.72	4.09*
Implementation of REDD	1,16	.24	Implementation of REDD	1,16	.29
International context	1,16	.33	International context	1,16	2.24
Co-benefits of REDD	1,16	.28	Co-benefits of REDD	1,16	.08
Temporal aspects[1]	1,14.09	.23	Temporal aspects	1,16	.276
Politics and negotiations	1,16	.09	Politics and negotiations	1,16	.00

[1] Welch-test (robust against violation of equal variances) reported as the assumption of homogeneity of variance was violated for ANOVA.

* p < .10, ** p < .05, *** p < .01

Table 31. Centrality and specificity scores on concept maps, students REDD exercise.

	Centrality				*Specificity*		
	Mean score map 1 (X_1)	Mean score map 2 (X_2)	Change (X_1-X_2)		Mean score map 1 (X_1)	Mean score map 2 (X_2)	Change (X_1-X_2)
Financing & sources	1.48	1.19	0.29	Financing and sources	0.1929	0.2047	-0.0118
Allocation/ disbursement	2.22	1.94	0.28	Allocation/ disbursement	0.1201	0.1278	-0.0077
Governance of mechanism	2.75	1.94	0.81	Governance of mechanism	0.1025	0.1200	-0.0176
Env. performance of REDD	2.00	1.91	0.09	Environmental performance of REDD	0.1597	0.1457	0.0140
Policy objectives	2.59	3.32	-0.72	Policy objectives	0.0807	0.0538	0.0269
Actors	2.90	3.42	-0.52	Actors	0.0727	0.0500	0.0227
Implementation of REDD	2.52	3.03	-0.51	Implementation of REDD	0.1009	0.0707	0.0302
International	3.94	4.10	-0.16	International	0.0119	0.0126	-0.0007

context				context			
Co-benefits of REDD	3.78	2.46	1.32	Co-benefits of REDD	0.0574	0.0954	-0.0380
Temporal aspects	3.45	3.00	0.45	Temporal aspects	0.0351	0.0433	-0.0082
Politics and negotiations	3.04	3.23	-0.19	Politics and negotiations	0.0660	0.0759	-0.0099

Table 32. Summary of analysis of variances. Differences of centrality and specificity scores between pre- and post-test, , students REDD exercise.

	Centrality			*Specificity*	
	df	F		df	F
Financing and sources[1]	1,130.99	3.73*	Financing and sources	1,136	1.52
Allocation/ disbursement	1,136	1.43	Allocation/ disbursement	1,136	.43
Governance of mechanism[1]	1,129.18	9.46***	Governance of mechanism	1,136	3.97**
Env. performance of REDD	1,136	.11	Env. performance of REDD[1]	1,127.17	1.37
Policy objectives	1,136	6.39**	Policy objectives[1]	1,114.03	5.94**
Actors	1,136	3.64*	Actors[1]	1,114.53	5.05**
Implementation of REDD	1,136	3.01*	Implementation of REDD[1]	1,112.79	8.50***
International context	1,136	.51	International context	1,136	.05
Co-benefits of REDD[1]	1,121.18	26.08***	Co-benefits of REDD[1]	1,96.48	27.41***
Temporal aspects[1]	1,132.06	3.35*	Temporal aspects	1,136	3.08
Politics and negotiations[1]	1,131.85	.43	Politics and negotiations	1,136	.87

[1] Welch-test (robust against violation of equal variances) reported as the assumption of homogeneity of variance was violated for ANOVA.
* p < .10, ** p < .05, *** p < .01

Table 33. Summary of analysis of variances. Between-group (expert/student) differences in centrality and specificity scores at pre-test, REDD exercise.

	Centrality			*Specificity*	
	df	F		df	F
Financing and sources[1]	1,8.81	.32	Financing and sources	1,76	1.19
Allocation/disbursement[1]	1,27.49	14.43**	Allocation/disbursement	1,76	7.30**
Governance of mechanism	1,76	.02	Governance of mechanism	1,76	.35

Environmental performance of REDD	1,76	3.83**	Environmental performance of REDD	1,76	2.49
Policy objectives	1,76	.85	Policy objectives	1,76	1.28
Actors	1,76	.00	Actors	1,76	.01
Implementation of REDD	1,76	.06	Implementation of REDD	1,76	.01
International context[1]	1,8.98	5.81**	International context[1]	1,8.11	8.51**
Co-benefits of REDD	1,76	.59	Co-benefits of REDD	1,76	.16
Temporal aspects	1,76	.51	Temporal aspects[1]	1,12.39	.26
Politics and negotiations[1]	1,20.02	18.52***	Politics and negotiations[1]	1,8.45	.42

[1] Welch-test (robust against violation of equal variances) reported as the assumption of homogeneity of variance was violated for ANOVA.

* p < .10, ** p < .05, *** p < .01

Table 34. Summary of analysis of variances. Between-group (expert/student) differences in centrality and specificity scores at post-test, REDD exercise.

	Centrality			*Specificity*	
	df	F		df	F
Financing and sources[1]	1,8.48	.42	Financing and sources	1,76	
Allocation/ disbursement	1,76	1.12	Allocation/ disbursement	1,76	
Governance of mechanism	1,76	.11	Governance of mechanism	1,76	
Environmental performance of REDD	1,76	3.58	Environmental performance of REDD	1,76	
Policy objectives	1,76	1.07	Policy objectives[1]	1,8.70	1.33
Actors[1]	1,12.52	3.19*	Actors[1]	1,41.30	9.11***
Implementation of REDD	1,76	.16	Implementation of REDD	1,76	
International context	1,76	6.05**	International context[1]	1,8.22	2.55
Co-benefits of REDD	1,76	4.53**	Co-benefits of REDD[1]	1,35.36	20.09***
Temporal aspects	1,76	.16	Temporal aspects[1]	1,9.12	.21
Politics and negotiations[1]	1,18.95	10.86***	Politics and negotiations[2]	1,76	2.32

[1] Welch-test (robust against violation of equal variances) reported as the assumption of homogeneity of variance was violated for ANOVA.

[2]Homogeneity of variance violated for ANOVA, but Welch-test could not be conducted as the variance in one group was 0.

* p < .10, ** p < .05, *** p < .01

Annex 2.C Additional evidence for cognitive learning from concept map analysis

Table 35. Summary of analysis of variance. Difference in number of items and levels per participant group between pre- and post-test.

			df	F
Experts	EU Burden-sharing	Items	1, 8	.29
		Levels	1, 8	.00
	REDD	Items	1, 16	.30
		Levels	1, 16	.35
Students	EU Burden-sharing	Items	1, 118	23.31***
		Levels	1, 118	3.25*
	REDD	Items	1, 136	13.16***
		Levels	1,136	4.11**

* p < .10, ** p < .05, *** p < .01

Table 36. Summary of analysis of variance. Between-group differences (experts vs. students; REDD vs. burden-sharing) at pre- and at post-test.

			df	F
Experts vs. students	No. of items	pre-test	1,141	3.68*
		post-test	1,141	.66
	No. of levels	pre-test	1,141	3.45*
		post-test	1,141	1.34
Burden-sharing vs. REDD	No. of items	pre-test	1,141	6.15**
		post-test	1,141	1.99
	No. of levels	pre-test	1,141	.63
		post-test	1,141	.12

* p < .10, ** p < .05, *** p < .01

Table 37. Shift in response types from pre- to post-concept maps, per participant group. Results from Wilcox Signed-Rank test.

		z
Experts	EU Burden-sharing	.00
	REDD	-1.73*
Students	EU Burden-sharing	-3.67***
	REDD	-2.74***

* p < .10, ** p < .05, *** p < .01

Annex 2.D Survey results, self-reported cognitive learning

Table 38. Proximity to policy-making and self-reported cognitive learning (experts).

		N	Mean	Standard Deviation
Cognitive learning about current policy	Close to policy-making (public sector/NGO)	7	2.7143	.95119
	Further removed from policy-making (research/private sector)	19	3.6842	.82007
	Total	26	3.4231	.94543
Cognitive learning about future policy	Close to policy-making (public sector/NGO)	7	3.2857	1.25357
	Further removed from policy-making (research/private sector)	19	3.6316	.95513
	Total	26	3.5385	1.02882
Cognitive learning about policy dynamics	Close to policy-making (public sector/NGO)	7	3.8571	.37796
	Further removed from policy-making (research/private sector)	19	3.7368	.93346
	Total	26	3.7692	.81524

Table 39. Academic background and self-reported cognitive learning (students, burden-sharing exercise).

		N	Mean	Standard Deviation
Cognitive learning about current policy	Political science/IR/European studies	11	3.7273	1.10371
	Social sciences and environmental subjects	25	3.8400	.80000
	Natural sciences/Humanities/others	11	3.1818	1.16775
	Total	47	3.6596	.98415
Cognitive learning about future policy	Political science/IR/European studies	11	4.2727	1.00905
	Social sciences and environmental subjects	25	4.2800	.84261
	Natural sciences/Humanities/others	11	4.5455	.68755
	Total	47	4.3404	.84124

Cognitive learning about policy dynamics	Political science/IR/European studies	11	3.9091	.94388
	Social sciences and environmental subjects	25	3.8400	1.06771
	Natural sciences/Humanities/others	11	4.0000	.77460
	Total	47	3.8936	.96084

Table 40. Pearson correlations, motivation of experts to participate in the policy exercise and self-reported cognitive learning.

	Cognitive learning about current policy	Cognitive learning about future policy	Cognitive learning about policy dynamics
To learn more about current developments in European climate policy/ REDD financing	.299	-.088	.321
To better understand the dynamics of EU climate policy-making/ international climate policy-making	.456*	-.105	.149
To learn more about future developments in European climate policy/REDD financing	.036	-.044	-.068
To better understand the different perspectives of stakeholders in European climate policy/ REDD	.269	.077	.491*
Interest in the work of the ADAM project/the REDD-ALERT project	.409	.062	.472*
Opportunity to network	.472*	.235	.583**
Policy exercise format	.525*	-.029	.433

* p < .05, ** p < .01

Annex 2.E Survey results, recorded normative change from pre- to post-test

Table 41. T-test results for pre- and post-survey statements for the expert run of the burden-sharing exercise

	Survey 1 Std. Error Mean (SE_1)	Survey 2 Std. Error Mean (SE_2)	t	Sig. (2-tailed)
In your view, how important are the following factors in reaching agreement on effort-sharing in the EU?				
Sound and transparent economic models with clear criteria underlying member state targets	0.23	0.14	-0.43	0.68
Arrangements which take account of historic emission reductions	0.16	0.19	1.00	0.34
Willingness of the economically most advanced member states to shoulder a comparatively larger mitigation burden	0.21	0.25	1.17	0.27
Willingness of the less advanced member states to make climate change a political priority, even if this comes at a certain economic cost	0.18	0.18	0.00	1.00
Compensation or 'side payments' in other policy domains	0.25	0.26	-1.00	0.34
Pressure for the EU to lead internationally	0.14	0.18	1.84	**0.10***
Strong role of the Commission as an impartial facilitator	0.33	0.18	-1.00	0.34
In your view, how important are the following factors for the successful implementation of an effort-sharing agreement?				
Fairness of targets	0.20	0.21	-1.00	0.34
An economic context which allows for reaching the targets at reasonable costs to the economy	0.23	0.18	0.00	1.00
A strict compliance regime, with heavy sanctions for non-compliers	0.19	0.18	-0.43	0.68
Temporal flexibility, including options for the banking and borrowing of emission reductions	0.25	0.27	-0.45	0.66
Fairness of allowance allocation under the EU ETS	0.23	0.19	0.00	1.00
Accurate and timely reporting of emission trends by member states	0.18	0.21	0.36	0.72

Good management on behalf of the Commission as regards monitoring and enforcement	0.16	0.19	-1.00	0.34
The push-effect of a global climate agreement including ambitious targets for all major emitters	0.21	0.19	1.61	0.14
Evidence of significant emission reductions globally	0.21	0.22	3.00	**0.01****
To what extent do you agree with the following statements in the context of burden-/effort-sharing in European climate policy?				
The EC protects the interests of the less developed countries too much	0.23	0.24	-2.63	**0.03****
The wealthier member states care too little about economic development of the poorer countries	0.28	0.28	0.00	1.00
The economically less advanced member states should attach a higher priority to climate policy in their national policies	0.16	0.21	.803	0.44
The Commission is pushing too hard to achieve an ambitious effort-sharing/target-setting outcome	0.35	0.28	-1.00	0.34

Table 42. Means and standard deviations in pre- and post-survey statements for the expert run of the burden-sharing exercise.

	Survey 1 Mean (X1)	Survey 2 Mean (X2)	Change (X2-X1)	Survey 1 Std. Deviation (SD1)	Survey 2 Std. Deviation (SD2)	Change (SD1-SD2)
In your view, how important are the following factors in reaching agreement on effort-sharing in the EU?						
Sound and transparent economic models with clear criteria underlying member state targets	4.18	4.27	0.09	0.75	0.54	0.21
Arrangements which take account of historic emission reductions	3.91	3.73	-0.18	0.47	0.65	-0.18
Willingness of the economically most advanced member states to shoulder a comparatively larger mitigation burden	4.45	4.09	-0.36	0.69	0.83	-0.14
Willingness of the less advanced member states to make climate change a political priority, even if this comes at a certain economic cost	4.18	4.18	0.00	0.60	0.60	0.00

Compensation or 'side payments' in other policy domains	3.55	3.82	0.27	0.82	0.87	-0.05
Pressure for the EU to lead internationally	4.27	3.82	-0.45	0.47	0.60	-0.14
Strong role of the Commission as an impartial facilitator	3.80	4.10	0.30	1.03	0.57	**0.47**
In your view, how important are the following factors for the successful implementation of an effort-sharing agreement?						
Fairness of targets	3.80	4.00	0.20	0.63	0.67	-0.03
An economic context which allows for reaching the targets at reasonable costs to the economy	4.18	4.18	0.00	0.75	0.60	0.15
A strict compliance regime, with heavy sanctions for non-compliers	3.73	3.82	0.09	0.65	0.60	0.04
Temporal flexibility, including options for the banking and borrowing of emission reductions	3.55	3.73	0.18	0.82	0.90	-0.08
Fairness of allowance allocation under the EU ETS	4.00	4.00	0.00	0.77	0.63	0.14
Accurate and timely reporting of emission trends by member states	4.18	4.09	-0.09	0.60	0.70	-0.10
Good management on behalf of the Commission as regards monitoring and enforcement	4.09	4.27	0.18	0.54	0.65	-0.11
The push-effect of a global climate agreement including ambitious targets for all major emitters	4.45	4.00	-0.45	0.69	0.63	0.06
Evidence of significant emission reductions globally	4.00	3.50	-0.50	0.67	0.71	-0.04
To what extent do you agree with the following statements in the context of burden-/effort-sharing in European climate policy?						
The EC protects the interests of the less developed countries too much	2.18	2.73	0.55	0.75	0.79	-0.04
The wealthier member states care too little about economic development of the poorer countries	3.09	3.09	0.00	0.94	0.94	0.00
The economically less advanced member states should attach a higher priority to climate policy in their national policies	4.10	3.90	0.20	0.54	0.70	0.16
The Commission is pushing too hard to achieve an ambitious effort-sharing/target-setting outcome	2.18	2.45	0.27	1.17	0.93	0.23

Table 43. T-test results for pre- and post-survey statements for the student run of the burden-sharing exercise

	Survey 1 Std. Error Mean (SE1)	Survey 2 Std. Error Mean (SE2)	t	Sig. (2-tailed)
In your view, how important are the following factors in reaching agreement on effort-sharing in the EU?				
Sound and transparent economic models with clear criteria underlying member state targets	0.09	0.12	1.61	.115
Arrangements which take account of historic emission reductions	0.15	0.13	-4.05	.000**
Willingness of the economically most advanced member states to shoulder a comparatively larger mitigation burden	0.13	0.14	-1.04	.303
Willingness of the less advanced member states to make climate change a political priority, even if this comes at a certain economic cost	0.12	0.09	-2.22	.032*
Compensation or 'side payments' in other policy domains	0.13	0.13	0.15	.878
Pressure for the EU to lead internationally	0.16	0.18	0.64	.523
Strong role of the Commission as an impartial facilitator	0.13	0.14	-1.96	.057
In your view, how important are the following factors for the successful implementation of an effort-sharing agreement?				
Fairness of targets	0.12	0.10	0.60	.555
An economic context which allows for reaching the targets at reasonable costs to the economy	0.13	0.11	-1.03	.309
A strict compliance regime, with heavy sanctions for non-compliers	0.13	0.12	0.17	.868
Temporal flexibility, including options for the banking and borrowing of emission reductions	0.14	0.15	-1.92	**.062**
Fairness of allowance allocation under the EU ETS	0.11	0.14	1.50	.141
Accurate and timely reporting of emission trends by member states	0.15	0.12	-0.62	.538
Good management on behalf of the Commission as regards monitoring and enforcement	0.11	0.10	1.43	.160

The push-effect of a global climate agreement including ambitious targets for all major emitters	0.13	0.13	-0.46	.645
Evidence of significant emission reductions globally	0.15	0.16	0.85	.403
To what extent do you agree with the following statements in the context of burden-/effort-sharing in European climate policy?				
The EC protects the interests of the less developed countries too much	0.12	0.12	-1.48	.146
The wealthier member states care too little about economic development of the poorer countries	0.15	0.18	0.67	.507
The economically less advanced member states should attach a higher priority to climate policy in their national policies	0.16	0.13	-0.60	.553
The Commission is pushing too hard to achieve an ambitious effort-sharing/target-setting outcome	0.13	0.13	-0.67	.507

Table 44. Means and standard deviations in pre- and post-survey statements for the student run of the burden-sharing exercise.

	Survey 1 Mean (X_1)	Survey 2 Mean (X_2)	Change (X_2-X_1)	Survey 1 Std. Deviation (SD_1)	Survey 2 Std. Deviation (SD_2)	Change (SD_1-SD_2)
In your view, how important are the following factors in reaching agreement on effort-sharing in the EU?						
Sound and transparent economic models with clear criteria underlying member state targets	4.20	3.95	-0.24	0.60	0.77	-0.17
Arrangements which take account of historic emission reductions	3.12	3.90	0.78	0.98	0.83	0.15
Willingness of the economically most advanced member states to shoulder a comparatively larger mitigation burden	4.00	4.12	0.12	0.84	0.87	-0.03
Willingness of the less advanced member states to make climate change a political priority, even if	3.76	4.02	0.27	0.77	0.57	0.20

this comes at a certain economic cost						
Compensation or 'side payments' in other policy domains	3.30	3.28	-0.02	0.82	0.82	0.01
Pressure for the EU to lead internationally	3.61	3.51	-0.10	1.02	1.16	-0.14
Strong role of the Commission as an impartial facilitator	3.37	3.68	0.32	0.86	0.91	-0.05
In your view, how important are the following factors for the successful implementation of an effort-sharing agreement?						
Fairness of targets	4.39	4.32	-0.07	0.77	0.61	0.16
An economic context which allows for reaching the targets at reasonable costs to the economy	4.00	4.15	0.15	0.84	0.69	0.15
A strict compliance regime, with heavy sanctions for non-compliers	4.07	4.05	-0.02	0.82	0.77	0.05
Temporal flexibility, including options for the banking and borrowing of emission reductions	3.33	3.65	0.33	0.89	0.92	-0.03
Fairness of allowance allocation under the EU ETS	4.26	4.03	-0.23	0.72	0.90	-0.19
Accurate and timely reporting of emission trends by member states	4.02	4.10	0.07	0.96	0.80	0.16
Good management on behalf of the Commission as regards monitoring and enforcement	4.32	4.17	-0.15	0.69	0.67	0.02
The push-effect of a global climate agreement including ambitious targets for all major emitters	3.80	3.88	0.07	0.84	0.84	0.00
Evidence of significant emission reductions globally	3.76	3.61	-0.15	0.97	1.05	-0.08
To what extent do you agree with the following statements in the context of burden-/effort-sharing in European climate policy?						
The EC protects the interests of the less developed countries too much	2.49	2.68	0.20	0.75	0.79	-0.04
The wealthier member states care too little about economic development of the poorer countries	2.93	2.80	-0.12	0.96	1.14	-0.19
The economically less advanced member states should attach a higher priority to climate policy in their national policies	3.39	3.49	0.10	1.02	0.84	0.18
The Commission is pushing too hard to achieve an ambitious effort-sharing/target-setting outcome	2.34	2.46	0.12	0.82	0.84	-0.01

Table 45. T-test results for pre- and post-survey statements of the expert run of the REDD exercise.

	Survey 1 Std. Error Mean (SE_1)	Survey 2 Std. Error Mean (SE_2)	t	Sig. (2-tailed)
A global REDD+ fund is a valuable interim step before linking REDD+ to the global carbon market.	0.41	0.20	-1.32	.225
Establishing the structures for a global REDD+ fund might lead to 'institutional lock-in' and delay or altogether prevent the integration of REDD+ credits into the global carbon market.	0.26	0.32	0.32	.760
A global performance-based REDD+ fund is the most adequate instrument for financing REDD+ also in the long term.	0.31	0.42	0.00	1.000
A global REDD+ fund will never generate a meaningful level of funding.	0.42	0.34	1.47	.179
A global performance-based REDD+ fund should make co-benefits, such as biodiversity, a key criterion for disbursing funds.	0.28	0.56	0.82	.438
A balanced regional distribution of funding should be of highest priority for a global REDD+ fund.	0.36	0.41	-0.45	.665
A global REDD+ fund should be managed by an institution with long-standing experience and expertise in the field, like the World Bank or the GEF.	0.26	0.29	0.22	.834
Establishing a REDD+ mechanism that will effectively reduce tropical deforestation rates is almost impossible	0.22	0.28	-1.84	**.104**

Table 46. Means and standard deviations in pre- and post-survey statements of the expert run of the REDD exercise.

	Survey 1 Mean (X_1)	Survey 2 Mean (X_2)	Change (X_2-X_1)	Survey 1 Std. Deviation (SD_1)	Survey 2 Std. Deviation (SD_2)	Change (SD_1-SD_2)
A global REDD+ fund is a valuable interim step before linking REDD+ to the global carbon market.	3.67	4.11	0.44	1.22	0.60	0.62

Establishing the structures for a global REDD+ fund might lead to 'institutional lock-in' and delay or altogether prevent the integration of REDD+ credits into the global carbon market.	2.89	2.78	-0.11	0.78	0.97	-0.19
A global performance-based REDD+ fund is the most adequate instrument for financing REDD+ also in the long term.	3.11	3.11	0.00	0.93	1.27	-0.34
A global REDD+ fund will never generate a meaningful level of funding.	3.11	2.56	-0.56	1.27	1.01	0.26
A global performance-based REDD+ fund should make co-benefits, such as biodiversity, a key criterion for disbursing funds.	3.78	3.44	-0.33	0.83	1.67	-0.83
A balanced regional distribution of funding should be of highest priority for a global REDD+ fund.	3.22	3.44	0.22	1.09	1.24	-0.14
A global REDD+ fund should be managed by an institution with long-standing experience and expertise in the field, like the World Bank or the GEF.	3.11	3.00	-0.11	0.78	0.87	-0.08
Establishing a REDD+ mechanism that will effectively reduce tropical deforestation rates is almost impossible.	1.78	2.22	0.44	0.67	0.83	-0.17

Table 47. T-test results for ranking exercise in pre- and post-surveys of the expert run of the REDD exercise.

	Survey 1 Std. Error Mean (SE_1)	Survey 2 Std. Error Mean (SE_2)	t	Sig. (2-tailed)
Disbursement criteria	0.19	0.24	0.63	0.551
Modalities for verification of performance	0.17	0.19	-1.18	0.277
Monitoring requirements	0.21	0.18	0.64	0.542
Fund capitalisation and replenishment	0.21	0.21	1.53	0.170
Accounting for the co-benefits of REDD+	0.22	0.21	0.73	0.487
Institutional arrangements	0.22	0.22	-0.19	0.857
Modalities for stakeholder participation	0.19	0.22	0.00	1.000

Table 48. Means and standard deviations for ranking exercise in pre- and post-surveys of the expert run of the REDD exercise.

	Survey 1 Mean (X_1)	Survey 2 Mean (X_2)	Change (X_2-X_1)	Survey 1 Std. Deviation (SD_1)	Survey 2 Std. Deviation (SD_2)	Change (SD_1-SD_2)
Disbursement criteria	3.75	3.13	-0.63	1.98	2.30	-0.31
Modalities for verification of performance	3.63	4.50	0.88	2.62	1.51	1.10
Monitoring requirements	4.50	4.00	-0.50	1.85	1.51	0.34
Fund capitalisation and replenishment	3.25	2.25	-1.00	1.67	2.05	-0.38
Accounting for the co-benefits of REDD+	4.50	4.00	-0.50	1.85	1.77	**0.08**
Institutional arrangements	3.63	3.75	0.13	2.07	1.83	0.23
Modalities for stakeholder participation	4.75	4.75	0.00	2.25	2.19	**0.06**

Table 49. T-test results (based on corrected data) for ranking exercise in pre- and post-surveys of the student runs of the REDD exercise.

	Survey 1 Std. Error Mean (SE_1)	Survey 2 Std. Error Mean (SE_2)	t	Sig. (2-tailed)
Disbursement criteria	0.19	0.24	-2.61	.011
Modalities for verification of performance	0.17	0.19	-0.79	.435
Monitoring requirements	0.21	0.18	0.81	.422
Fund capitalisation and replenishment	0.21	0.21	0.36	.716
Accounting for the co-benefits of REDD+	0.22	0.21	0.90	.369
Institutional arrangements	0.22	0.22	2.75	.008
Modalities for stakeholder participation	0.19	0.22	-0.92	.362

Table 50. Means and standard deviations for ranking exercise in pre- and post-surveys of the student runs of the REDD exercise (based on corrected data).

	Survey 1 Mean (X_1)	Survey 2 Mean (X_2)	Change (X_2-X_1)	Survey 1 Std. Deviation (SD_1)	Survey 2 Std. Deviation (SD_2)	Change (SD_1-SD_2)
Disbursement criteria	2.57	3.38	0.80	1.58	2.02	-0.43

Modalities for verification of performance	3.97	4.16	0.19	1.42	1.60	-0.17
Monitoring requirements	4.50	4.29	-0.21	1.74	1.44	0.30
Fund capitalisation and replenishment	2.35	2.27	-0.08	1.76	1.73	0.03
Accounting for the co-benefits of REDD+	4.33	4.12	-0.20	1.78	1.74	0.04
Institutional arrangements	4.52	3.71	-0.82	1.83	1.79	0.04
Modalities for stakeholder participation	5.35	5.60	0.25	1.59	1.83	-0.24

Table 51. T-test results for pre- and post-survey statements of the student runs of the REDD exercise.

	Survey 1 Std. Error Mean (SE_1)	Survey 2 Std. Error Mean (SE_2)	t	Sig. (2-tailed)
A global REDD+ fund is a valuable interim step before linking REDD+ to the global carbon market.	0.08	0.12	0.70	.489
Establishing the structures for a global REDD+ fund might lead to 'institutional lock-in' and delay or altogether prevent the integration of REDD+ credits into the global carbon market.	0.11	0.12	-1.75	.084
A global performance-based REDD+ fund is the most adequate instrument for financing REDD+ also in the long term.	0.14	0.14	-0.09	.928
A global REDD+ fund will never generate a meaningful level of funding.	0.11	0.12	-0.18	.857
A global performance-based REDD+ fund should make co-benefits, such as biodiversity, a key criterion for disbursing funds.	0.12	0.14	2.29	.025
A balanced regional distribution of funding should be of highest priority for a global REDD+ fund.	0.13	0.13	1.28	.205
A global REDD+ fund should be managed by an institution with long-standing experience and expertise in the field, like the World Bank or the GEF.	0.12	0.15	0.61	.544
Establishing a REDD+ mechanism that will effectively reduce tropical deforestation rates is almost impossible	0.11	0.13	-0.90	.374

Table 52. Means and standard deviations in pre- and post-survey statements of the student runs of the REDD exercise.

	Survey 1 Mean (X_1)	Survey 2 Mean (X_2)	Change (X_2-X_1)	Survey 1 Std. Deviation (SD_1)	Survey 2 Std. Deviation (SD_2)	Change (SD_1-SD_2)
A global REDD+ fund is a valuable interim step before linking REDD+ to the global carbon market.	4.25	4.17	-0.08	0.70	1.01	-0.31
Establishing the structures for a global REDD+ fund might lead to 'institutional lock-in' and delay or altogether prevent the integration of REDD+ credits into the global carbon market.	2.83	3.08	0.25	0.95	1.06	-0.11
A global performance-based REDD+ fund is the most adequate instrument for financing REDD+ also in the long term.	3.47	3.48	0.01	1.19	1.20	-0.01
A global REDD+ fund will never generate a meaningful level of funding.	2.69	2.72	0.03	0.97	1.06	-0.09
A global performance-based REDD+ fund should make co-benefits, such as biodiversity, a key criterion for disbursing funds.	4.04	3.72	-0.32	1.04	1.23	-0.19
A balanced regional distribution of funding should be of highest priority for a global REDD+ fund.	3.24	3.04	-0.20	1.11	1.17	-0.05
A global REDD+ fund should be managed by an institution with long-standing experience and expertise in the field, like the World Bank or the GEF.	3.43	3.33	-0.09	1.05	1.27	-0.21
Establishing a REDD+ mechanism that will effectively reduce tropical deforestation rates is almost impossible	2.27	2.36	0.09	0.98	1.11	-0.12

Annex 2.F Method for recoding responses for the ranking exercise in the REDD survey

We applied this method to all response sets where more than one item had received the same rank(s), or where one or two items had not received a score at all (responses with more than two missing scores were excluded from the analysis). In recoding the responses, we sought to retain the relative position of each item in relation to the others, while at the same time adjusting its value in such a way that the mean for all seven items, after recoding, would equal that of a correctly ranked set of items (M(1,2,3,4,5,6,7)=4). For instance, for a set where two items were ranked '1', we recoded both with the score of 1.5, to reflect that they jointly occupied the first and second out of seven ranks. For illustration, a few examples from the reare included below.

Item	A	B	C	C	E	F	G	Mean value
Original response set	2	2	3	3	4	5	6	3.57
Response set after recoding	**1.5**	**1.5**	**3.5**	**3.5**	**5**	**6**	**7**	**4**
Original response set	1	1	1	4	5	5	6	3.29
Response set after recoding	**2**	**2**	**2**	**4**	**5.5**	**5.5**	**7**	**4**
Original response set	2	2	3	3	7	7	7	4.43
Response set after recoding	**1.5**	**1.5**	**3.5**	**3.5**	**6**	**6**	**6**	**4**
Original response set	1	2	3	4	5	6		3.5
Response set after recoding	**1**	**2.2**	**3.4**	**4.6**	**5.8**	**7**		**4**

References

Ackoff, R. L. (1974). *Redesigning the future. A systems approach to societal problems*. New York: John Wiley & Sons.

Acton, W. H., Johnson, P. J., & Goldsmith, T. E. (1994). Structural knowledge assessment: comparison of referent structures. *Journal of Educational Psychology*, *86*(2), 303–311.

Adger, W. N., Arnell, N. W., & Tompkins, E. L. (2005). Adapting to climate change: perspectives across scales. *Global Environmental Change*, *15*(2), 75–76.

Agrawal, A., Nepstad, D., & Chhatre, A. (2011). Reducing emissions from deforestation and forest degradation. *Annual Review of Environment and Resources*, *36*(1), 373–396.

All, A. C., & Havens, R. L. (1997). Cognitive/concept mapping: a teaching strategy for nursing. *Journal of Advanced Nursing*, *25*(6), 1210–1219.

Allwood, J. M., Bosetti, V., Dubash, N., Gómez-Echeverri, L., & von Stechow, C. (2014). *Working group III - mitigation of climate change. Annex I: glossary*. Geneva.

Angelsen, A., Brown, S., Loisel, C., Peskett, P., Streck, C., & Zarin, D. (2009). *Reducing emissions from deforestation and forest degradation (REDD): An options assessment report*. Washington D.C.: Meridian Institute.

Argyris, C. (2003). A life full of learning. *Organization Studies*, *24*(7), 1178–1192.

Argyris, C., & Schön, D. A. (1980). *Organizational learning*. Reading, MA: Addison-Wesley.

Argyris, C., & Schön, D. A. (1996). *Organizational learning II: Theory, method and practice*. Reading, MA: Addison-Wesley.

Armitage, D. R., Berkes, F., & Doubleday, N. (2007). *Adaptive co-management. Collaboration, learning, and multi-level governance*. Vancouver: UBC Press.

Armitage, D. R., Marschke, M., & Plummer, R. (2008). Adaptive co-management and the paradox of learning. *Global Environmental Change*, *18*(1), 86–98.

Armitage, D. R., Plummer, R., Berkes, F., Arthur, R. I., Charles, A. T., Davidson-Hunt, I. J., Diduck, A.P., Doubleday, N. C., Johnson, D. S., Marschke, M., McConney, P., Pinkerton, E. W. & Wollenberg, E. K. (2009). Adaptive co-management for social–ecological complexity. *Frontiers in Ecology and the Environment*, *7*(2), 95–102.

Ausubel, D. P. (1962). A subsumption theory of meaningful verbal learning and retention. *The Journal of General Psychology*, *66*(2), 213–224.

Axelrod, R., & Cohen, M. D. (1999). *Harnessing complexity: organizational implications of a scientific frontier*. New York: Free Press.

Bachofen, C., Suarez, P., Steenbergen, M., & Grist, N. (2013). *Can games help people manage the climate risks they face? The participatory design of educational games.* Geneva: Red Cross/Red Crescent Climate Centre.

Baird, J., Plummer, R., Haug, C., & Huitema, D. (2014). Learning effects of interactive decision-making processes for climate change adaptation. *Global Environmental Change*, *27*, 51–63.

Bandura, A. (1977). *Social learning theory*. New York: General Learning Press.

Bekebrede, G. (2010). *Experiencing complexity. A gaming approach for understanding infrastructure systems*. Delft: Next Generation Infrastructures Foundation.

Bennett, C. J., & Howlett, M. (1992). The lessons of learning - reconciling theories of policy learning and policy change. *Policy Sciences*, *25*(3), 275–294.

Blackmore, C. (2007). What kinds of knowledge, knowing and learning are required for addressing resource dilemmas? A theoretical overview. *Environmental Science & Policy*, *10*(6), 512–525.

Blackstock, K. L., Kelly, G. J. J., & Horsey, B. L. L. (2007). Developing and applying a framework to evaluate participatory research for sustainability. *Ecological Economics*, *60*(4), 726–742.

Bots, P., & van Daalen, E. (2007). Functional design of games to support natural resource management policy development. *Simulation & Gaming*, *38*(4), 512–532.

Brewer, G. D. (1986). Methods for synthesis: policy exercises. In W. C. Clark & R. E. Munn (Eds.), *Sustainable development of the biosphere* (pp. 455–473). Cambridge: Cambridge University Press.

Brown, H. S., & Vergragt, P. (2008). Bounded socio-technical experiments as agents of systemic change: the case of a zero-energy residential building. *Technological Forecasting and Social Change*, *75*(1), 107–130.

Bryman, A., & Cramer, D. (1990). *Quantitative data analysis for social scientists*. London: Routledge.

Burke, S., & Parthemore, C. (2009). *Climate change war game: major findings and background*. Washington D.C.: Center for a New American Security.

Burkhalter, S., Gastil, J., & Kelshaw, T. (2002). A conceptual definition and theoretical model of public deliberation in small face-to-face groups. *Communication Theory*, *12*(4), 398–422.

Busenberg, G. J. (2001). Learning in organizations and public Policy. *Journal of Public Policy*, *21*(02), 173–189.

Buzan, T., & Buzan, B. (1993). *The mind map book*. London: BBC Books.

Caffrey, M. (2000). Toward a history-based doctrine for wargaming. *Aerospace Journal*, *14*(3), 33–56.

Cañas, A., Coffey, J. W., Carnot, M. J., Feltovich, J., Novak, J. D., & Hoffman, R. R. (2003). *A summary of literature pertaining to the use of concept mapping techniques and technologies for education and performance support*. Pensacola, FL: The Institute for Human and Machine Recognition.

Checkel, J. T. (2001). Why comply? Social learning and European identity change. *International Organization*, *55*(3), 553–588.

Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The Nature of expertise*. Hillsdale, NJ: L. Erlbaum Associates.

Clark, W. C. (1986). Sustainable development of the biosphere: themes for a research program. In W. C. Clark & R. E. Munn (Eds.), *Sustainable development of the biosphere (*pp. 5–48). Cambridge: Cambridge University Press.

Cline, B. E., Brewster, C. C., & Fell, R. D. (2010). A rule-based system for automatically evaluating student concept maps. *Expert Systems with Applications*, *37*(3), 2282–2291.

Corbera, E., Estrada, M., & Brown, K. (2010). Reducing greenhouse gas emissions from deforestation and forest degradation in developing countries: revisiting the assumptions. *Climatic Change*, *100*(3), 355–388.

Craps, M. (ed. . (2003). *Social learning in river basin management*. Leuven: Centre for Organizational and Personnel Psychology, K.U. Leuven.

Crona, B. I., & Parker, J. N. (2012). Learning in support of governance: Theories, methods, and a framework to assess how bridging organizations contribute to adaptive resource governance. *Ecology and Society*, *17*(1), 32.

Crookall, D., & Thorngate, W. (2009). Acting, knowing, learning, simulating, gaming. *Simulation and Gaming*, *40*(1), 8–26.

Cuppen, E. (2009). *Putting perspectives into participation*. Oisterwijk: Uitgeverij BOXPress.

Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm*. Englewood Cliffs, NJ: Prentice-Hall.

Czaja, R., & Blair, J. (2005). *Designing surveys: A guide to decisions and procedures. 2nd edition*. Thousand Oaks, CA: Pine Forge Press.

Daniels, S. E., & Walker, G. B. (1996). Collaborative learning: Improving public deliberation in ecosystem-based management. *Environmental Impact Assessment Review, 16*(2), 71–102.

De Caluwé, L. I. A. (1997). *Veranderen moet je leren.* Den Haag: Delwel Publishers.

Derbentseva, N., Safayeni, F., Can, J., & Cañas, A. (2007). Concept maps: Experiments on dynamic thinking. *Journal of Research in Science Teaching*, *44*(3), 448–465.

Deutsch, M. (1973). *The resolution of conflict. Constructive and destructive processes*. New Haven, CT: Yale University Press.

Deyle, R., & Schively Slotterback, C. (2009). Group learning in participatory planning processes. *Journal of Planning Education and Research*, *29*(1), 23–38.

Diduck, A. P. (2010). The learning dimension of adaptive capacity: Untangling the multi-level connections. In D. R. Armitage & R. Plummer (Eds.), *Adaptive Capacity and Environmental Governance* (pp. 199–221). Berlin, Heidelberg: Springer.

Dielemann, H., & Huisingh, D. (2006). Games by which to learn and teach about sustainable development: Exploring the relevance of games and experiential learning for sustainability. *Journal of Cleaner Production*, *14*(9), 837-847.

Draper, S. W. (2006). *The Hawthorne, Pygmalion, placebo and other effects of expectation: some notes*. unpublished manuscript, available at http://www.psy.gla.ac.uk/steve/hawth.html (last accessed 14.04.2010).

Dryzek, J. S. (2000). *Deliberative democracy and beyond. Liberals, critics, contestations*. Oxford: Oxford University Press.

Duinker, P. N., Nilsson, S., & Toth, F. L. (1993). *Testing the "policy exercise" in studies of Europe's s forest sector: methodological reflections on a bittersweet experience*. Laxenburg: Institute for Applied Systems Analysis (IIASA).

Duke, R. D. (1974). *Gaming: The future's language*. London: Sage.

Duke, R. D. (2011). Origin and evolution of policy simulation: A personal journey. *Simulation & Gaming*, *42*(3), 342–358.

Duke, R. D., & Geurts, J. (2004). *Policy games for strategic management*. Amsterdam: Dutch University Press.

Easton, D. (1953). *The political system: An inquiry into the state of political science*. New York: Knopf.

Easton, D. (1965). *A systems analysis of political life*. New York: John Wiley & Sons.

Eden, C. (2004). Analyzing cognitive maps to help structure issues or problems. *European Journal of Operational Research*, *159*(3), 673–686.

Eiser, J. R. (1986). *Social psychology. Attitudes, cognition, and social behaviour*. Cambridge: Cambridge University Press.

Etheredge, L. S. (1981). Government learning: An overview. In S. L. Long (Ed.), *The Handbook of political behavior, vol.2* (pp. 73–161). New York: Plenum Press.

Etheredge, L. S., & Short, J. (1983). Thinking about government learning. *Journal of Management Studies*, *20*(1), 41–58.

European Commission. (2014). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A policy framework for climate and energy in the period from 2020 to 2030*. COM (2014)15. Brussels: European Commission.

Folke, C., Hahn, T., Olsson, P., & Norberg, J. (2005). Adaptive governance of social-ecological systems. *Annual Review of Environment and Resources*, *30*(1), 441–473.

Forester, J. (1999). *The deliberative practitioner. Encouraging participatory planning processes*. Cambridge, MA: MIT Press.

Foucault, M. (1981). The order of discourse. In R. Young (Ed.), *Untying the Text: A Post-structuralist Reader* (pp. 48–79). London: Routledge, Kegan and Paul.

Freeman, R. (2006). Learning in public policy. In M. Moran, M. Rein, & R. E. Goodin (Eds.), *The Oxford handbook of public policy* (pp. 367–388). Oxford: Oxford University Press.

Funtowicz, S. O., & Ravetz, J. R. (1991). A new scientific methodology for global environmental issues. In R. Costanza (Ed.), *Ecological Economics. The Science and Management of Sustainability* (pp. 137–152). New York: Columbia University Press.

Funtowicz, S. O., & Ravetz, J. R. (1993). Science for the post-normal age. *Futures*, *25*(7), 739–755.

Garmendia, E., & Stagl, S. (2010). Public participation for sustainability and social learning: Concepts and lessons from three case studies in Europe. *Ecological Economics*, *69*(8), 1712–1722.

Geurts, J. L. A., Duke, R. D., & Vermeulen, P. A. M. (2007). Policy gaming for strategy and change. *Long Range Planning*, *40*(6), 535–558.

Geurts, J. L. A., & Vennix, J. (1989). *Verkenningen in beleidsanalyse*. Zeist: Kerckebosch.

Gibbons, M., Limoges, C., Nowotny, H., Trow, M., Scott, P., & Schwartzman, S. (1994). *The new production of knowledge*. London: Sage.

Gieryn, T. F. (1983). Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American Sociological Review*, *48*(6), 781–795.

Glasbergen, P. (1996). Learning to manage the environment. In J. Meadowcraft (Ed.), *Democracy and the environment: Problems and prospects* (pp. 175–193). Cheltenham: Edward Elgar.

Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology*, *40*, 631–666.

Greenblat, C. S., & Duke, R. D. (1975). *Gaming-simulation: Rationale, design, and applications.* Beverly Hills, CA: Sage.

Gregersen, H., Lakany, H. El, White, A., El Lakany, H., & Karsenty, A. (2010). *Does the opportunity cost approach indicate the real cost of REDD+?* Washington, DC: Rights and Resources Initiative.

Gregory, R., Ohlson, D., & Arvai, J. (2006). Deconstructing adaptive management: Criteria for applications to environmental management. *Ecological Applications*, *16*(6), 2411–2425.

Grin, J., & Loeber, A. (2006). Theories of policy learning: Agency, structure, and change. In F. Fischer, G. J. Miller, & M. S. Sydney (Eds.), *Handbook of Public Policy Analysis* (pp. 201–219). London: Taylor & Francis.

Groves, D. G., Knopman, D., Lempert, R., Berry, S. H., & Wainfan, L. (2008). *Presenting uncertainty about climate change to water-resource managers. A summary of workshops with the inland empire utilities agency*. Santa Monica, CA: RAND Corporation.

Grubb, M. (2013). Doha's dawn? *Climate Policy*, *13*(3), 281–284.

Gutmann, A., & Thompson, D. Y. (1996). *Democracy and disagreement*. Cambridge, MA: Harvard University Press.

Habermas, J. (1984). *The theory of communicative action*. Boston, MA: Beacon Press.

Hajer, M. A. (1995). *The politics of environmental discourse. Ecological modernization and the policy process*. Oxford: Clarendon Press.

Hall, P. A. (1993). Policy paradigms, social learning, and the state: The case of economic policymaking in Britain. *Comparative Politics*, *25*(3), 275–296.

Harteveld, C. (2012). *Making sense of virtual risks. A quasi-experimental investigation into game-based training*. Amsterdam, Fairfax: IOS Press.

Haug, C., & Berkhout, F. (2010). Learning the hard way. European climate policy after Copenhagen. *Environment*, *52*(3), 20–27.

Haug, C., & Gupta, J. (2013). REDD and the global climate regime. In J. Gupta, N. van de Grijp, & O. Kuik (Eds.), *Climate Change, forests and REDD: Lessons for institutional design.* Abingdon: Routledge.

Haug, C., Huitema, D., & Wenzler, I. (2011). Learning through games? Evaluating the learning effect of a policy exercise on European climate policy. *Technological Forecasting and Social Change*, *78*(6), 968–981.

Haug, C., & Jordan, A. (2010). Burden-sharing: Distributing burdens or sharing efforts. In A. Jordan, D. Huitema, H. van Asselt, F. Berkhout, & T. Rayner (Eds.), *Climate change policy in the European Union: Confronting the dilemmas of adaptation and mitigation?* (pp. 83–103). Cambridge: Cambridge University Press.

Haug, C., & Pattberg, P. H. (2008). Der REDD-Hype in den Klimaverhandlungen. Internationale Waldschutzdebatte nach Bali. *Informationsbrief Weltwirtschaft Und Entwicklung (W&E)*, *19*(9).

Haug, C., Rayner, T., Jordan, A., Hildingsson, R., Stripple, J., Monni, S., Huitema, D., Massey, E., van Asselt,. H., & Berkhout, F. (2010). Navigating the dilemmas of climate policy in Europe: Evidence from policy evaluation studies. *Climatic Change*, *101*(3-4), 427–445.

Heclo, H. (1974). *Modern social politics in Britain and Sweden. From relief to income maintenance*. New Haven, CT: Yale University Press.

Hedberg, B. (1981). How organizations learn and unlearn. In P. C. Nystrom & W. H. Starbuck (Eds.), *Handbook of organizational design* (pp. 3–27). Oxford: Oxford University Press.

Hickson, D. J. (1986). *Top decisions. Strategic decision-making in organizations*. London: Basil Blackwell.

Hill, T., & Lewicki, P. (2006). *Statistics: methods and applications*. Tulsa, OK: Statsoft.

Hisschemöller, M., Tol, R. S. J., & Vellinga, P. (2001). The relevance of participatory approaches in integrated environmental assessment. *Integrated Assessment*, *2*(2), 57–72.

Holling, C. S. (1978). *Adaptive environmental assessment and management*. Chichester: John Wiley & Sons.

Hoppe, R. (2010). *The governance of problems. Puzzling, powering, participation*. Bristol: The Policy Press.

Howlett, M. (2011). *Designing public policies. Principles and instruments*. Abingdon: Routledge.

Huitema, D., Cornelisse, C., & Ottow, B. (2010). Is the jury still out? Toward greater insight in policy learning in participatory decision processes-the case of Dutch citizens' juries on water management in the Rhine Basin. *Ecology and Society*, *15*(1), 16.

Huitema, D., Jordan, A., Massey, E., Rayner, T., van Asselt, H., Haug, C., Hildingsson, R., Monni, S., Stripple, J. (2011). The evaluation of climate policy: Theory and emerging practice in Europe. *Policy Sciences*, *44*(2), 179–198.

Huitema, D., van de Kerkhof, M., & Pesch, U. (2007). The nature of the beast: Are citizens ' juries deliberative or pluralist? *Policy Sciences*, *40*(4), 287–311.

Humphreys, D. (2008). The politics of "avoided deforestation": Historical context and contemporary issues. *International Forestry Review*, *10*(3), 433–442.

Jacobs-Lawson, J. M., & Hershey, D. A. (2002). Concept maps as an assessment tool in psychology courses. *Teaching of Psychology*, *29*(1), 25–29.

Jäger, J., Sonntag, N., & Bernard, D. (1990). *The challenge of sustainable development in a greenhouse world: Some visions for the future. Report of a policy exercise in Bad Bleiberg, Austria, September 2-7, 1990*. Stockholm: Stockholm Environment Institute.

Jann, W., & Wegrich, K. (2007). Theories of the policy cycle. In F. Fischer, G. Miller, & M. Sidney (Eds.), *Handbook of public policy analysis: Theory, politics and methods* (pp. 43–58). London: Taylor & Francis.

Jarosz, W., & Nye, J. (1993). Shadow of the past: Learning from history in national security decision-making. In P. Tetlock, J. Husbands, R. Jervis, P. Stern, & C. Tilly (Eds.), *Behavior, Society and International Conflict, vol. III* (pp. 126–189). New York: Oxford University Press.

Jasanoff, S. (1990). *The fifth branch. Science advisers as policymakers*. Cambridge, MA: Harvard University Press.

Joldersma, C. (2000). Policy learning through simulation/gaming. In D. Saunders & N. Smalley (Eds.), *Simulations and games for transition and change* (pp. 77–85). London: Kogan Page.

Joldersma, C., Geurts, J. L. A., & van't Spijker, W. J. H. (1995). Spelsimulatie. Theorie, definitie en plaatsbepaling in de bestuurskunde. *Bestuurskunde*, *4*(4), 148–156.

Joldersma, C., & Roelofs, A. M. E. (2004). The impact of soft OR-methods on problem structuring. *European Journal of Operational Research*, *152*(3), 696–708.

Jones, A. (2009). *The Copenhagen climate exercise - a simulation-based role playing exercise*. Hartland,VT: Sustainability Institute.

Jones, W. M. (1964). Crisis games for adults and others. In R. Levine, T. Schelling & W. Jones (Eds.), *Crisis games 27 years later: Plus c'est déjà vu* (pp. 39–46). Santa Monica, CA: RAND Corporation.

Jordan, A., Huitema, D., & van Asselt, H. (2010). Climate change policies in the European Union. An introduction. In A. Jordan, D. Huitema, H. van Asselt, T. Rayner, & F. Berkhout (Eds.), *Climate change policy in the european union: Confronting the dilemmas of adaptation and mitigation?* (pp. 3–28). Cambridge: Cambridge University Press.

Jordan, A., van Asselt, H., Berkhout, F., Huitema, D., & Rayner, T. (2012). Understanding the paradoxes of multilevel governing : Climate change policy in the European Union. *Global Environmental Politics*, *12*(2), 43–66.

Kaiser, F. G., & Fuhrer, U. (2003). Ecological behavior's dependency on different forms of knowledge. *Applied Psychology - an International Review*, *52*(4), 598–613.

Kasemir, B., Toth, F. L., & Masing, V. (2003). Venture capital and climate policy. In B. Kasemir, J. Jäger, C. Jaeger, & M. T. Gardner (Eds.), *Public participation in sustainability science: A handbook* (pp. 155–175). Cambridge: Cambridge University Press.

Kay, J. J., & Schneider, E. (1994). Embracing complexity: The challenge of the ecosystem approach. *Alternatives*, *20*(3), 32–38.

Keen, M., Brown, V., & Dybal, R. (2005). *Social learning in environmental management*. London: Earthscan.

Kinchin, I. M., Hay, D. B., & Adams, A. (2000). How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. *Educational Research*, *42*(1), 43–57.

King, C., & Jiggins, J. (2002). A systematic model and theory for facilitating social learning. In C. Leeuwis & A. Pyburn (Eds.), *Wheelbarrows full of frogs. Social learning in rural resource management* (pp. 85–105). Assen: Koninklijke Van Gorcum.

Kingdon, J. W. (1984). *Agendas, alternatives, and public policies*. *Second edition*. Don Mills, ON: HarperCollins.

Klabbers, J. H. G. (2006). *The magic circle. Principles of gaming & simulation*. Rotterdam: Sense Publishers.

Klabbers, J. H. G. (2009). Terminological ambiguity: Game and simulation. *Simulation & Gaming*, *40*(4), 446–463.

Klabbers, J. H. G., Swart, R. J., van Ulden, A. P., & Vellinga, P. (1995). Management of organized complexity through gaming. In D. Crookall & K. Arai (Eds.), *Simulation and gaming across disciplines: ISAGA at a watershed* (pp. 122–134). Thousand Oaks, CA: Sage.

Kloprogge, P., & Sluijs, J. P. Van Der. (2006). The inclusion of stakeholder knowledge and perspectives in integrated assessment of climate change. *Climatic Change*, *75*(3), 359–389.

Kolb, D. A. (1984). *Experiential learning*. Englewood Cliffs, NJ.: Prentice Hall.

Kriz, W. C. (2003). Creating effective learning environments and learning organizations through gaming simulation design. *Simulation & Gaming*, *34*(4), 495–511.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Lave, J., & Wenger, E. (1991). *Situated learning: legitimate peripheral participation*. Cambridge: Cambridge University Press.

Laws, D., & Rein, M. (2003). Reframing practice. In M. A. Hajer & H. Wagenaar (Eds.), *Deliberative policy analysis. Understanding governance in the network society* (pp. 172–206). Cambridge: Cambridge University Press.

Lederman, L. C. (1984). Debriefing: A critical reexamination of the postexperience analytic process with implications for its effective use. *Simulation & Gaming*, *15*(4), 415–431.

Lederman, L. C. (1992). Debriefing: Toward a Systematic Assessment of Theory and Practice. *Simulation & Gaming*, *23*(2), 145–160.

Lee, K. N. (1993). *Compass and gyroscope: integrating science and politics for the environment*. Washington, DC: Island Press.

Lee, K. N. (1999). Appraising adaptive management. *Ecology and Society*, *3*(2).

Leeuwis, C., & Pyburn, R. (2002). *Wheelbarrows full of frogs: Social learning in rural resource management*. Assen: Van Gorcum.

Levin, K., Cashore, B., Bernstein, S., & Auld, G. (2012). Overcoming the tragedy of super wicked problems: Constraining our future selves to ameliorate global climate change. *Policy Sciences*, *45*(2), 123–152.

Levitt, B., & March, J. G. (1988). Organizational learning. *Annual Review of Sociology*, *14*(1988), 319–340.

Levy, J. S. (1994). Learning and foreign policy : Sweeping a conceptual minefield. *International Organization*, *48*(2), 279–312.

Lewicki, R. J., Saunders, D. M., & Minton, J. W. (1997). *Essentials of negotiation*. Chicago, IL: Irwin.

Lindblom, C. E. (1959). The science of "muddling through." *Public Administration Review*, *19*(2), 79–88.

Lindblom, C. E. (1968). *The policy-making process*. Englewood Cliffs, NJ: Prentice-Hall.

Lomask, M., Baron, J. B., Greig, J., & Harrison, C. (1992). *ConnMap: Connecticut's use of concept mapping to assess the structure of students' knowledge of science*. Paper presented at the Annual Meeting of the National Association of Research in Science Teaching. Cambridge, MA.

Lonsdale, K. G., Downing, T. E., Nicholls, R. J., Parker, D., Vafeidis, A. T., Dawson, R., & Hall, J. (2008). Plausible responses to the threat of rapid sea-level rise in the Thames Estuary. *Climatic Change*, *91*(1), 145–169.

Luckie, D., Harrison, S. H., & Ebert-May, D. (2011). Model-based reasoning: using visual tools to reveal student learning. *Advances in Physiology Education*, *35*(1), 59–67.

Maarleveld, M., & Dangbegnon, C. (2002). Social Learning: Major concepts and issues. In C. Leeuwis & A. Pyburn (Eds.), *Wheelbarrows full of frogs: Social learning in rural resource management* (pp. 67–84). Assen: Van Gorcum.

Mason, R. O., & Mitroff, I. I. (1981). *Challenging strategic planning assumptions: Theory, cases, and techniques*. New York: John Wiley & Sons.

Massey, E., Biesbroek, R., Huitema, D., & Jordan, A. (2014). The adoption and diffusion of climate change adaptation policies across Europe. *Global Environmental Change*, (29), 434–443.

Massey, E., & Huitema, D. (2012). The emergence of climate change adaptation as a policy field: the case of England. *Regional Environmental Change*, *13*(2), 341–352.

Mathews, F. D. (1994). *Politics for people. Finding a responsible public voice*. Urbana, IL: University of Illinois Press.

May, P. J. (1992). Policy learning and failure. *Journal of Public Policy*, *12*(4), 331.

Mayer, I. S., Bekebrede, G., Harteveld, C., Warmelink, H., Zhou, Q., Van Ruijven, T., Lo, J., Kortmann, R., & Wenzler, I. (2014). The research and evaluation of serious games: Toward a comprehensive methodology. *British Journal of Educational Technology*, *45*(3), 502–527.

Mayer, I. S. (2009). The gaming of policy and the politics of gaming: A Review. *Simulation & Gaming*, *6*(4), 825–862.

Mayer, I. S., & Veeneman, W. (2002). *Games in a world of infrastructures. Simulation-games for research, learning and intervention*. Delft Netherlands: Eburon.

McClure, J. R., Sonak, B., & Suen, H. K. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of Research in Science Teaching*, *36*(4), 475–492.

Mendler de Suarez, J., Suarez, P., Bachofen, C., Fortugno, N., Goentzel, J., Goncalves, P., Grist, N., Macklin, C., Pfeifer, K., Schweizer, S., van Aalst, M., & Virji, H. (2012). *Games for a new climate: Experiencing the complexity of future risks*. Boston: The Frederick S. Pardee Center for the Study of the Longer-Range Future, Boston University.

Mezirow, J. (1995). Transformation theory in adult learning. In M. R. Welton (Ed.), *In defense of the life world* (pp. 39–70). Albany, NY: State University of New York Press.

Mezirow, J. (1996). Contemporary paradigms of learning. *Adult Education Quarterly*, *46*(3), 158–172.

Milbrath, L. W. (1989). *Envisioning a sustainable society. Learning our way out*. Albany, NY: State University of New York Press.

Morine-Dershimer, G. (1993). Tracing conceptual change in preservice teachers. *Teaching and Teacher Education*, *9*(1), 15–26.

Mouton, J., & Marais, H. C. (1988). *Basic concepts in the methodology of the social sciences*. Pretoria: Human Sciences Research Council.

Munaretto, S., & Huitema, D. (2012). Adaptive co-management in the Venice lagoon? An analysis of current water and environmental management practices and prospects for for change. *Ecology and Society*, *17*(2), 19.

Muro, M., & Jeffrey, P. (2008). A critical review of the theory and application of social learning in participatory natural resource management processes. *Journal of Environmental Planning and Management*, *51*(3), 325–344.

Muro, M., & Jeffrey, P. (2012). Time to talk? How the structure of dialog processes shapes stakeholder learning in participatory water resource management. *Ecology And Society*, *17*(1), 3.

National Research Council. (2009). *Informing decisions in a changing climate. Panel on strategies and methods for climate-related decision support*. Washington, DC: The National Academies Press.

Nemeth, C., Brown, K., & Rogers, J. (2001). Devil's advocate versus authentic dissent: Stimulating quantity and quality. *European Journal of Social Psychology*, *31*(6), 707–720.

Newig, J., & Pahl-Wostl, C. (2010). Synapses in the network: Learning in governance networks in the context of environmental management. *Ecology and So*ciety, *15*(4).

Nilsson, A. E., & Gerger Swartling, A. (2009). *Social learning about climate adaptation: Global and local perspectives*. Stockholm: Stockholm Environment Institute.

Nilsson, M. (2005). The role of assessments and institutions for policy learning: A study on Swedish climate and nuclear policy formation. *Policy Sciences*, *38*(4), 225–249.

Novak, J. D. (1998). *Learning, creating, and using knowledge. Concept maps as facilitative tools in schools and corporations*. Mahwah, NL: Erlbaum Associates.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge: Cambridge University Press.

Obersteiner, M., Hüttner, M., Kraxner, F., McCallum, I., Aoki, K., Böttcher, H., Fritz, S., Gusti, M., Havlik, P., & Kindermann, G. (2009). On fair, effective and efficient REDD mechanism design. *Carbon Balance and Management*, *4*(11).

Oberthür, S., & Pallemaerts, M. (2010). *The new climate policies of the European Union: Internal legislation and climate diplomacy*. Brussels: VUB Press.

OECD. (2001). *The Well-being of nations. The role of human and social capital*. Paris: Organisation for Economic Cooperation and Development.

Ogonowski, H., Helme, N., Movius, D., & Schmidt, J. (2007). *Reducing emissions from deforestation and degradation: The dual markets approach*. Washington, DC: Center for Clean Air Policy.

Ostrom, E., & Ahn, T. K. (2003). *Foundations of social capital*. Cheltenham: Edward Elgar.

Owens, S., Rayner, T., & Bina, O. (2004). New agendas for appraisal: Reflections on theory, practice, and research. *Environment and Planning A*, *36*(11), 1943–1959.

Pahl-Wostl, C. (2006). The importance of social learning in restoring the multifunctionality of rivers and floodplains. *Ecology and Society*, *11*(1), 10.

Pahl-Wostl, C. (2009). A conceptual framework for analysing adaptive capacity and multi-level learning processes in resource governance regimes. *Global Environmental Change*, *19*(3), 354–365.

Pahl-Wostl, C., Craps, M., Dewulf, A., Mostert, E., Tabara, D., & Taillieu, T. (2007). Social learning and water resources management. *Ecology and Society*, *12*(2).

Parson, E. A. (1995). *Global climate change policy exercise: Results of a test run*. Laxenburg: International Institute for Applied Systems Analysis.

Parson, E. A. (1996a). Three dilemmas in the integrated assessment of climate change. *Climatic Change*, *34*(3), 315–326.

Parson, E. A. (1996b). What can you learn from a game? In R. J. Zeckhauser, R. L. Keeney, & J. K. Sebenius (Eds.), *Wise choices. Decisions, games and negotiations* (pp. 233–252). Boston, MA: Harvard Business School Press.

Parson, E. A. (1997). Informing global environmental policy-making: A plea for new methods of assessment and synthesis. *Environmental Modelling and Assessment*, *2*(4), 267–279.

Parson, E. A., & Clark, W. C. (1995). Sustainable development as social learning: Theoretical perspectives and practical challenges for the design of a research program. In L. Gunderson, C. S. Holling, & S. Light (Eds.), *Barriers*

and bridges to the renewal of ecosystems and institutions (pp. 428–460). New York: Columbia University Press.

Parson, E. A., & Ward, H. (1998). Games and simulations. In S. Rayner & E. L. Malone (Eds.), *Human Choice and Climate Change* (Vol. 3, pp. 105–140). Columbus, OH: Battelle Press.

Parsons, W. (1995). *Public policy: An introduction to the theory and practice of policy analysis*. Cheltenham: Edward Elgar.

Patton, M. Q. (1990). *Qualitative evaluation and research methods. 2nd edition*. Newbury Park, CA: Sage.

Pearce, W. B., & Littlejohn, S. W. (1997). *Moral conflict. When social worlds collide*. Thousand Oaks, CA: Sage.

Pearsall, N. R., Skipper, J. O. E. L. J., & Mintzes, J. J. (1996). Knowledge restructuring in the life sciences: A longitudinal study of conceptual change in biology. *Science Education, 81*(2), 193–215.

Plummer, R., Crona, B., Armitage, D. R., Olsson, P., & Yudina, O. (2012). Adaptive co-management: A systematic review and analysis. *Ecology and Society*, *17*(3), 11.

Plummer, R., & FitzGibbon, J. (2007). Connecting adaptive co-management, social learning, and social capital through theory and practice. In D. Armitage, F. Berkes, & N. Doubleday (Eds.), *Adaptive co-management: collaboration, learning and multilevel governance* (pp. 38–61). Vancouver, BC: UBC Press.

Press, D. (1994). *Democratic dilemmas in the age of ecology. Trees and toxics in the American West*. Durham: Duke University Press.

Putnam, L. L., & Holmer, M. (1992). Framing, reframing, and issue development. In L. L. Putnam & M. E. Roloff (Eds.), *Communication and negotiation* (Vol. 20, pp. 128–155). Newbury Park: Sage.

Rajamani, L. (2012). The Durban platform for enhanced action and the future of the climate regime. *International and Comparative Law Quarterly*, *61*(02), 501–518.

Reed, M. S., Evely, A. C., Cundill, G., Fazey, I. R. A., Glass, J., Laing, A., Newig, J., Parrish, B., Prell, C., Raymond, C., & Stringer, L. C. (2010). What is social learning? Ecology and Society, 15(4), R1.

Renn, O., Webler, T., & Wiedemann, P. (1995). *Fairness and competence in citizen participation*. Dordrecht: Kluwer.

Rice, D. C., Ryan, J. M., & Samson, S. M. (1998). Using concept maps to assess student learning in the science classroom: Must different methods compete? *Journal of Research in Science Teaching*, *35*(10), 1103–1127.

Ringius, L. (1999). Differentiation, leaders, and fairness: Negotiating climate commitments in the European community. *International Negotiation*, *4*, 133–166.

Rist, S., Chiddambaranathan, M., Escobar, C., & Wiesmann, U. (2006). "It was hard to come to mutual understanding..." - the multidimensionality of social learning processes concerned with sustainable natural resource use in India, Africa and Latin America. *Systemic Practice and Action Research*, *19*(3), 219–237.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*(2), 155–169.

Rodela, R., Cundill, G., & Wals, A. E. J. (2012). An analysis of the methodological underpinnings of social learning research in natural resource management. *Ecological Economics, 77*, 16-26.

Roehler, L. R., Duffy, G. G., Herrmann, B. A., Conley, M., & Johnson, J. (1988). Knowledge structures as evidence of the "personal": Bridging the gap from thought to practice. *Journal of Curriculum Studies*, *20*(2), 159–165.

Roelofs, A. M. E. (2000). *Structuring policy issues. Testing a mapping technique with gaming/ simulation*. Tilburg: Van Spaendonck.

Rohrbaugh, J., & Eden, C. (1990). Using the competing values approach to explore "ways of working." In C. Eden & J. Radford (Eds.), *Tackling strategic problems: The role of group decision support* (pp. 40–47). London: Sage.

Rotmans, J. (1998). Methods for IA: The challenges and opportunities ahead. *Environmental Modeling and Assessment*, *3*(3), 155–179.

Rotmans, J., & van Asselt, M. B. A. (2003). Integrated assessment modelling. In P. Martens, J. Rotmans, D. Jansen, & K. Vrieze (Eds.), *Climate change: An integrated perspective* (pp. 239–275). Dordrecht: Springer.

Ruane, J. M. (2005). *Essentials of research methods: A guide to social research*. Malden, London, Victoria: Blackwell.

Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching*, *33*(6), 569-600.

Ryan, T. (2000). The role of simulation gaming in policy-making. *Systems Research and Behavioral Science*, *17*(4), 359–364.

Rye, J. A., & Rubba, P. A. (2002). Scoring concept maps: An expert map-based scheme weighted for relationships. *School Science and Mathematics*, *102*(1), 33–44.

Sabatier, P. A. (1987). Knowledge, policy-oriented learning, and policy change. *Science Communication*, *8*(4), 649–692.

Sabatier, P. A. (1988). An advocacy coalition framework of policy change and the role of policy-oriented learning therein. *Policy Sciences*, *21*(2-3), 129–168.

Sabatier, P. A. (1998). The advocacy coalition framework: Revisions and relevance for Europe. *Journal of European Public Policy*, *5*(1), 98–130.

Sabatier, P. A. (2007). The need for better theories. In P. A. Sabatier (Ed.), *Theories of the policy process. 2nd edition* (pp. 3–21). Boulder, CO: Westview Press.

Salen, K., & Zimmerman, E. (2006). *The game design reader: A rules of play anthology*. Cambridge, MA: MIT Press.

Salter, J., Robinson, J., & Wiek, A. (2010). Participatory methods of integrated assessment. A review. *Wiley Interdisciplinary Reviews: Climate Change*, *1*(5), 697–717.

Schahn, J. (1996). *Die Erfassung und Veränderung des Umweltbewusstseins*. Frankfurt/Main: Lang.

Schelling, T. C. (1964). An uninhibited sales pitch for crisis games. In R. Levine, T. Schelling & W. Jones (Eds.), *Crisis games 27 years later: Plus c'est déjà vu* (pp. 22–38). Santa Monica, CA: RAND Corporation.

Schively, C. (2007). A quantitative analysis of consensus building in local environmental review. *Journal of Planning Education and Research*, *27*(1), 82–98.

Schön, D. A. (1973). *Beyond the stable state. Public and private learning in a changing society*. Harmondsworth: Penguin.

Schön, D. A., & Rein, M. (1994). *Frame reflection: Toward the resolution of intractable policy controversies*. New York: Basic Books.

Schusler, T. M., Decker, D. J., & Pfeffer, M. J. (2003). Social learning for collaborative natural resource management. *Society & Natural Resources: An International Journal*, *16*(4), 309–326.

Senge, P. M. (1990). *The fifth discipline*. New York: Doubleday.

Stavins, R., Zou, J., Brewer, T., Conte Grand, M., den Elzen, M., Finus, M., Gupta, J., Höhne, N., Lee, M.-K., Michaelowa, A., Paterson, M., Ramakrishna, K., Wen, G., Wiener, J. & Winkler, H.G. (2014). International cooperation: Agreements and instruments. In O. Edenhofer, R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel, & J.C. Minx (Eds.), *Climate change 2014: Mitigation of climate change. contribution of Working Group II to the fifth assessment of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.

Stebbins, R. A. (2001). *Exploratory research in the social sciences*. Thousand Oaks, CA: Sage.

Stein, J. G. (1994). Political learning by doing: Gorbachev as uncommitted thinker and motivated learner. *International Organization*, *48*(2), 155–183.

Stern, N. (2006). *Stern review: The economics of climate change*. London: HM Treasury.

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., & Midgley, P.M. et al. (2013). *Climate change 2013. The physical science basis. Working group I contribution to the fifth assessment report of the Intergovernmental Panel on Climate Change. Abstract for decision-makers.* Cambridge: Cambridge University Press.

Stockwell, C., Hare, W., & Macey, K. (2009). Designing a REDD mechanism: the TDERM triptych. In B. J. Richardson, Y. Le Bouthillier, H. McLeod-Kilmurray, & S. Wood (Eds.), *Climate law and developing countries: Legal and policy challenges for the world economy* (pp. 151–177). Cheltenham: Edward Elgar.

Streck, C., Gomez-Echeverri, L., Gutman, P., Loisel, C., & Werksman, J. (2009). *REDD + institutional options assessment. Developing an effective, efficient, and equitable institutional framework for REDD+ under the UNFCCC*. Washington, DC: Meridian Institute.

Sunderlin, W. D., & Atmadja, S. (2009). Is REDD+ an idea whose time has come or gone. In A. Angelsen (Ed.) *Realising REDD+. National strategy and policy options* (pp. 45–53). Bogor: Center for International Forestry Research.

Surel, Y. (2000). The role of cognitive and normative frames in policy-making. *Journal of European Public Policy*, *7*(4), 495–512.

Tábara, J. D., & Pahl-Wostl, C. (2007). Sustainability learning in natural resource use and management. *Ecology and Society*, *12*(2), 3.

Tetlock, P. (1991). Learning in U.S. and Soviet foreign policy: In search of an elusive concept. In G. W. Breslauer & P. Tetlock (Eds.), *Learning in US and Soviet foreign policy* (pp. 20–61). Boulder, CO: Westview Press.

Tippett, J., Searle, B., Pahl-Wostl, C., & Rees, Y. (2005). Social learning in public participation in river basin management—early findings from HarmoniCOP European case studies. *Environmental Science & Policy*, *8*(3), 287–299.

Toth, F. L. (1988a). Policy exercises: Objectives and design elements. *Simulation & Gaming*, *19*(3), 235–255.

Toth, F. L. (1988b). Policy exercises: Procedures and implementation. *Simulation & Gaming*, *19*(3), 256–276.

Toth, F. L. (1992). Policy responses to climate change in Southeast Asia. In J. Schmandt & J. Clarkson (Eds.), *The regions and global warming: Impacts and response strategies* (pp. 304–322). New York: Oxford University Press.

Toth, F. L. (2003). State of the art and future challenges for integrated environmental assessment. *Integrated Assessment*, *4*(4), 250–264.

Toth, F. L., & Hizsnyik, E. (1998). Integrated environmental assessment methods: Evolution and applications. *Environmental Modeling and Assessment*, *3*(3), 193–207.

Toth, F. L., & Hizsnyik, E. (2008). Managing the inconceivable: Participatory assessments of impacts and responses to extreme climate change. *Climatic Change*, *91*(1), 81–101.

Turnbull, N. J. (2005). *Policy in question: From problem solving to problematology*. Sydney: Dissertation, University of New South Wales.

Turnpenny, J., Radaelli, C. M., Jordan, A., & Jacob, K. (2009). The policy and politics of policy appraisal: Emerging trends and new directions. *Journal of European Public Policy*, *16*(4), 640–653.

Ulrich, M. (1997). Games/Simulations about environmental issues. Existing tools and underlying concepts. In J. L. A. Geurts, C. Joldersma, & E. Roelofs (Eds.), *Gaming/simulation for policy development and organizational change* (pp. 303–311). Tilburg: Tilburg University Press.

Underwood, S. E., & Duke, R. D. (1987). Decisions at the top: Gaming as an aid to formulating policy options. In D. Crookall, C. Greenblat, A. Coote, J. Klabbers, & D. Watson (Eds.), *Simulation-gaming in the late 1980s* (pp. 289–296). Oxford: Pergamon Press.

United Nations Framework Convention on Climate Change. (2013). *Decision 9/CP.19: Work programme on results-based finance to progress the full implementation of the activities referred to in decision 1/CP.16, paragraph 70*. FCCC/CP/2013/10/Add.1. UNFCCC: Bonn.

Van Asselt, H., & Zelli, F. (2013). Connect the dots: managing the fragmentation of global climate governance. *Environmental Economics and Policy Studies*, *16*(2), 137-155.

Van Asselt, M. B. A., & Rijkens-Klomp, N. (2002). A look in the mirror: Reflection on participation in integrated assessment from a methodological perspective. *Global Environmental Change*, *12*(3), 167–184.

Van de Kerkhof, M. (2004). *Debating Climate Change*. Utrecht: Lemma.

Van de Kerkhof, M. (2006). Making a difference: On the constraints of consensus building and the relevance of deliberation in stakeholder dialogues. *Policy Sciences*, *39*(3), 279–299.

Van der Meer, F. B., & Geurts, J. L. A. (1995). Simulatie voor beleids-en organisatie-ontwikkeling. *Bestuurskunde*, *4*(4), 166–177.

Visser, M. (2007). Deutero-learning in organizations: A review and a reformulation. *Academy of Management Review*, *32*(2), 659–667.

Vissers, G. A. N., Heyne, G., & Peters, V. A. M. (1995). Spelsimulatie in bestuurskundig onderzoek. *Bestuurskunde*, *4*(4), 178–187.

Walters, C. J., & Holling, C. S. (1990). Large-scale management experiments and learning by doing. *Ecology*, *71*(6), 2060-2068.

Warren, M. E. (1996). What should we expect from more democracy? Radical democratic responses to politics. *Political Theory*, *24*, 241–270.

Weber, M. (2007). *From Max Weber. Essays in Sociology*. Hoboken: Taylor & Francis.

Webler, T., Kastenholz, H., & Renn, O. (1995). Public participation in impact assessment: A social learning perspective. *Environmental Impact Assessment Review*, *15*(5), 443–463.

Weiss, C. (1977). Research for policy's sake: The enlightenment function of social research. *Policy Analysis*, *3*(4), 531–545.

Wenzler, I. (1993). *Policy exercises: A new approach to policy development*. Nijmegen: ITS.

Williams, B. A., & Matheny, R. A. (1995). *Democracy, dialogue, and environmental disputes: The contested languages of social regulation*. New Haven: Yale University Press.

Yin, R. K. (2003). *Case Study Research Design and Methods. Third Edition*. Thousand Oaks, CA: Sage.

Yin, Y., Vanides, J., Ruiz-Primo, M. A., Ayala, C. C., & Shavelson, R. J. (2005). Comparison of two concept-mapping techniques: implications for scoring, interpretation and use. *Journal of Research in Science Teaching*, *42*(2), 166–184.

Zhou, Q. (2014). *The Princess in the castle*. Challenging Serious Game Play for Integrated Policy Analysis and Planning. Delft: Next Generation Infrastructures Foundation.

Summary

The urgency and scale of the climate problem require innovative ways of thinking about policy options that effectively address the complexities and scope of the challenge. Learning by policy-makers, scientists, and the public alike is needed to come to grips with the dilemmas posed by the specific characteristics of the climate challenge. The present thesis examines the potential offered by policy exercises, a type of simulation game involving policy-makers and experts, for policy-oriented learning. In doing so, it seeks to contribute to a better understanding of learning in and for environmental governance, at both the conceptual and the measurement level.

The notion of learning in and for environmental governance builds on several bodies of literature, of which the policy sciences, organisational theory, and natural resources and adaptive management are considered most relevant to the present work. Analyses of learning span a range of social units of analysis, from the level of the individual to policy-making systems at large. There are different views in the literature as to whether learning needs to result in an observable behavioural or policy change, and whether only certain types of change qualify as learning. This thesis defines policy-oriented learning as a change in understanding, which in the first instance occurs at the level of the individual. Contrary to many typologies which embrace a hierarchical understanding of learning, it argues that differentiating between different types of learning may be a more appropriate way of conceptualizing learning in an environmental governance context.

The conceptualisation of learning developed here distinguishes between cognitive, normative and relational learning. Cognitive learning is defined as the acquisition of new knowledge and facts, or the better structuring of existing knowledge, enabling more holistic or systems thinking. Normative learning refers to shifts in views, values or paradigms, which may vary in their degree of abstraction and social embeddedness. Relational learning finally involves a better understanding of others' mind sets, the development of trust and/or a better ability to cooperate. Applications of the typology have proven that it lends itself well to cross-case comparison, also when involving different evaluation teams. Distinguishing phenomena along the cognitive-normative-relational axis is intuitive. Moreover, the lack of a hierarchy between different levels of learning may reduce evaluator bias as there is no pressure to detect 'higher order' effects, which may not be forthcoming, especially if the appraisal process is rather short.

Policy exercises were first proposed in the late 1980s as a way to make sense of complex environmental problems and learn about potential ways of addressing them. They responded to the need of adjusting model-based simulation games, which had originally developed in a military context and over time broadened to other types of applications, to the challenges of long-term policy development in socio-ecological systems. Typically, participants in a policy exercise assume roles and take decisions in a structured setting that represents a simplified model of the system of reference that the game is supposed to represent. There are many claims in the literature as to the great learning potential that policy games and exercises offer to participants, from improved knowledge integration to stimulating creativity for developing novel policy options, including normative and relational benefits. Yet to date this potential – and the limitations of the method – have been hardly evaluated in a systematic fashion. This finding rings true for a broader range of interactive appraisal methods and represents the research gap that the present thesis seeks to address.

Despite their increasing application in environmental governance contexts, systematic evaluations of the learning effects of these methods are still lacking. Most efforts undertaken thus far fall short on several counts. Many do not sufficiently distinguish between different types of learning, but above all studies rarely involve a sufficient set of measurement tools and controls to allow for robust findings. The present thesis sets a higher conceptual and methodological standard in this respect. It devises a comprehensive evaluation framework that captures different types of learning and makes use of a variety of measurement instruments (concept maps, surveys, interviews, participant observations, collective debriefing) employed at multiple points in time. The framework thereby complies with standards stipulated in the evaluation literature, which underscore that assessments of learning should combine multiple methods and data sources and measure learning longitudinally.

We designed and ran two policy exercises on subjects in the realm of international climate policy, and evaluated the learning effects from these using the evaluation framework referred to above. The first exercise was about challenges related distributing emission reduction commitments among member states in the European Union, a process commonly known as ‘burden-sharing’ or ‘effort sharing’. The second exercise examined options to set up a global fund for REDD (Reducing Emissions from Deforestation and Forest Degradation) under the international climate regime. Such a fund would offer financial compensation to developing countries that succeed in protecting their forests, thereby avoiding emissions from lost forest carbon stocks.

Both exercises thus treat complex, unresolved puzzles that have a strong multi-actor, equity dimension. Both were run twice, once with international experts and once with Master’s students. This multiple-case research design allowed for comparing learning effects across the different groups, by disaggregating and aggregating the resulting evaluation data along two axes: the maturity of the topic of the exercise (EU

burden-sharing being a 'mature' policy topic, whereas REDD financing was more 'novel') and the level of expertise of the participants. In terms of limitations of the assessment, the thesis focused on participants' learning, rather than a comprehensive evaluation of the policy exercise. In terms of data and research design, the lack of a fully equivalent control group and the limited response rates for the post-exercise surveys and concept maps in the expert groups represent the most important constraints.

The results of the evaluation demonstrate that policy exercises can produce substantial learning effects that can be empirically measured and documented. Comparing the three types of learning over the different exercise runs, evidence was strongest for participants' cognitive learning, defined as the acquisition of new and the improved structuring of existing knowledge. Results from post-surveys and interviews indicate that participants gained insights both with regard to current policy developments and future policy options. More than anticipated given experts' experience in policy-making, policy consultancy or applied research, they also learned about the dynamics of policy-making and negotiations. This underscores the value of experiential learning for developing a deep understanding of process dynamics – whether in climate policy or in a different field. Analysis of pre- and post-concept maps also testified to changes in the emphasis of key concepts associated with the topic of the policy exercises and to improved knowledge structuring as evident from an increase in the average number of map levels and items.

The extent of normative and relational learning that we could ascertain was by comparison more limited. Survey results provided limited evidence for objectively 'recorded' normative change and hardly any indication of normative convergence within the group. Participants self-evaluated the extent of their normative learning more positively, although rather in terms of revised viewpoints on specific issues than larger-scale normative shifts. Findings also support the hypothesis that role switches can support and facilitate normative learning. Relational learning benefits from the exercises concerned primarily a better understanding of others' mind sets, positions and preferences. More specifically, participants reported insights into the interests of countries represented in the policy exercises and their underlying interests and motivations, as well as the wider actor configuration.

Comparing between-group effects, results confirmed the hypothesis that, on average, students learned more in the exercises than the experts. This was likely a function of their much less developed background knowledge on the policy context and the specific topic of the policy exercises. Yet our research also found significant evidence of learning in the expert group. This is remarkable given the short duration of the policy exercise and the high baseline from which the experts started out. The expectation that the greater 'novelty' of the issue of REDD financing compared to EU burden-sharing might enable higher cognitive and normative learning effects among

experts (for students, both topics were equally unfamiliar) was generally confirmed. This leads to the tentative conclusion that policy exercises hold greater potential for more unstructured, less mature topic areas.

The results of the evaluation confirm some, but not all of the benefits typically ascribed to policy exercises. Overall, the learning effects identified resonate with the emphasis on cognitive learning benefits voiced in the theoretical literature. There was evidence for improved knowledge integration and structuring, and participants self-reported insights into the advantages and drawbacks of various policy options. A broader, more holistic view of the problem equally transpired from some post-exercise survey responses and interviews. Indications of an enhanced long-term orientation and a serious reflection of the risks and contingencies involved in future policy-making were more ambiguous, however. Similarly, the policy innovation observed – genuinely new and innovative insights and policy options raised in the context of the exercises – was fairly limited. Ultimately, it may do more justice to the potential of the method to view the outputs from policy games and exercises as one step in a longer appraisal process rather than as an end in themselves.

On the upside, the exercises created a very interactive atmosphere that stimulated policy-makers and experts to share their knowledge and experience and engage in a lively dialogue over the proposed policy options, their broader impacts and associated challenges. The 'decision pressure' in the policy exercise played an important role in fostering this effective exchange across disciplinary and professional boundaries. This indicates that there is significant promise in using policy exercises as a tool to stimulate cognitive and relational learning at the interface of science and policy-making. Overall, a collaborative game design process involving key stakeholders from the start, a longer duration of the exercise as well as the 'right' set of participant, some of whom should ideally possess executive experience in policy-making, should further enhance the effectiveness of the method – though all require significant convening power on the part of the organisers.

In sum, the thesis broke new ground in systematically measuring the learning effects of policy exercises in the context of climate policy. Yet it concludes that there is still a substantial research agenda to be taken up for better understanding this tool and harnessing its potential for policy learning. First, there is still considerable scope in refining methods of measuring learning. The success of this thesis in demonstrating normative learning was limited. Here, new proxies for such learning and methods for measuring it are called for, perhaps linked to the analysis of concept maps. It is, however, also conceivable that normative learning as a concept may have less traction than anticipated, at least when it comes to policy appraisals among experts. Improving the measurement of learning is closely related to the second item on the research agenda: a better understanding of how the different facets of learning empirically relate to each other. In order to design better targeted interventions, we

need to understand the conditions under which different learning types reinforce or compete with each other. When it comes to unleashing the supposed creative potential of policy exercises for identifying new policy pathways, incrementalism will not do. We need a more fundamental game changer. Can policy exercises really live up to their most important function, and if so, under what circumstances? In many ways this represents the 'holy grail' of the profession, the aspect that is simultaneously most important and least advanced. In any case, and whatever the role that policy exercises can play in this, science is unequivocal in telling us that we need to make progress in environmental if we are to preserve our planet for future generations to enjoy.